

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

THIS PAGE BLANK (USPTO)

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: _____	NA Sequence (#) _____	STN _____
Searcher Phone #: If Contact: Sheppard	AA Sequence (#) _____	Dialog _____
Searcher Location: tel: 308-4499	Structure (#) _____	Questel/Orbi: _____
Date Searcher Picked Up: _____	Bibliographic _____	Dr.Link _____
Date Completed: 8/11/01	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems _____
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: _____	Other _____	Other (specify) _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 01:25:08 ; Search time 6114.61 Seconds
(without alignments)
3149.091 Million cell updates/sec

Title: US-09-402-713A-1

Perfect score: 2037

Sequence: 1 aagaagctgcatcagaaaaa.....caataaagaatttacaaga 2037

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

COMMENT

Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbs@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: 17
Class: BAC ends.

FEATURES

Location/Qualifiers
1..654
/organism="Homo sapiens"
/db_xref="GDB:754135"
/db_xref="taxon:9606"
/clone="RPCI-11-108L4"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"

BASE COUNT 213 a 100 c 139 g 202 t
ORIGIN

Query Match 30.1%; Score 613.8; DB 227; Length 654;
Best Local Similarity 98.5%; Pred. No. 1.2e-153;
Matches 645; Conservative 4; Mismatches 3; Indels 3; Gaps 3;

QY 1195 tgccattctcaagacctcaaatgtcattccatttaataatcacagatttaattttt 1254
DB 654 TGCCCATTCCTCAAGACCTCAAAATGTCATTCCATTAAATATCACAGGATTAACTTTT 595
QY 1255 ttaacctgagaagaattcaatgttaccatgcagctatgggaatttaataattttgttt 1314
DB 594 TTAACCTGGGAAGAAATCAATGTTACATGACGCTATGGGAATTAATATCATATTTTGT 535
QY 1315 tccagtgcagaatgactgaagtcctttatccctccctttgtttgtttttttccagta 1374
DB 534 TCCAGTGCAAGATGACTAAGTCTTTATCCCTCCCTTTGTTGATTTTTCAGTA 475
QY 1375 taaagttaaaatgcttagccttgactgaggtgtatatacag-cacagcctctcccatcc 1433
DB 474 TAAAGTTAAATGCTTAGCCTTTGACTGAGGCTGTATACAGGCACACAGCCTCTCCCATCC 415
QY 1434 cttccagccttatctgtcatcaccatcaacccctcccatnysaccctaaacaaatcctaa 1493
DB 414 CTCACGCTTATCTGTCTATCACCATCAACCCCTCCCATG-CACCTAAACAAATCTAACT 356
QY 1494 tgtaattccttgaaatgctcaggncatatactttctctgcctgagaagctcttccct 1553
DB 355 TGTAAATCTTGAACATGTCAGG-CATACATTAATCTCTCTGCTGAGGAGCTCTCTCT 297
QY 1554 gtctcttaantctagaatgatgtaaagtttgaaatgtaagttaattcttacttcatgcaa 1613
DB 296 GTCTCTTAAATCTAGAAATGATGTTAAAGTTTGAATAAGTTGACTATCTTACTTCATGCAA 237
QY 1614 agaaaggacacatatagagattcatcatcatcagacagacaaataactaaagttaattt 1673
DB 236 AGAAGGGACACATATGAGATTTCATCATCATGAGACACAAATACATAAAGTGAATTT 177
QY 1674 gattataagagtttagataaataatataatgaatgaacagkccacagaggaagtgttatg 1733
DB 176 GATTATAAGAGTTTAGATAAATATATGAATGCAAGAGCCACAGAGGGAATGTTTATGGG 117
QY 1734 gcacgtttgtaagcctggaggtgtaagaaagcaggggaacctcatagtattatataa 1793
DB 116 GCACGTTTCTAAGCCTGGGATGTGTAAGCAAGGACGAGGAACTCATAGTATCTTATATAA 57

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	513.8	30.1	654	227	AQ319247	RPC11-11-108L4
c 2	512.2	25.1	856	21	A1557225	PT2.1.15
c 3	399.6	19.6	441	170	BF858890	RC5-F019
c 4	394	19.3	402	9	AA578773	nh24a04.s
c 5	362	17.8	394	225	AQ206972	HS-3238.B
c 6	306.2	15.0	332	170	BF858286	RC5-F019
c 7	260.4	12.8	290	147	BF373619	MR0-F017
c 8	226.8	11.1	282	147	BF373581	MR0-F017
c 9	214.4	10.5	226	170	BF858371	RC5-F019
c 10	167	8.2	167	147	BF373406	IL2-F015
c 11	138.2	6.8	451	226	AQ246715	HS-2059.B
c 12	86.4	4.2	657	21	A1557495	PT2.1.7.H
c 13	63	3.1	1101	219	CNS0039G	
c 14	56	2.7	1001	219	CNS01400	
c 15	52.6	2.6	1101	219	CNS0182P	
c 16	52.2	2.6	1101	219	CNS0100X	
c 17	51.6	2.5	1101	219	CNS0000D1	
c 18	49.2	2.4	1204	219	CNS016E2	
c 19	48.2	2.4	1101	219	CNS017KX	
c 20	47.8	2.3	1101	219	CNS0039G	
c 21	47.6	2.3	529	225	AQ173559	HS-3202.A
c 22	47.6	2.3	1101	219	CNS0039E	
c 23	47.6	2.3	1101	219	CNS000EVL	
c 24	47.6	2.3	1101	219	CNS0106X	
c 25	47.2	2.3	1201	219	CNS0160R	
c 26	46.8	2.3	654	223	AQ046642	RPC11-35
c 27	45.2	2.2	1086	219	CNS00YXX	
c 28	44.8	2.2	938	219	CNS006TJ	
c 29	44.8	2.2	943	219	CNS002IM	
c 30	44.8	2.2	1125	106	AL547503	
c 31	44.4	2.2	926	219	CNS008LM	
c 32	44.4	2.2	1101	219	CNS017KT	
c 33	43.6	2.1	1101	219	CNS00LT2	
c 34	43.2	2.1	352	12	AA791841	vs65bll.r
c 35	43.2	2.1	656	119	AW691330	NF043E12S
c 36	43.2	2.1	997	219	CNS005TE	
c 37	43.2	2.1	1101	219	CNS017KE	
c 38	43	2.1	928	219	CNS0076D	
c 39	43	2.1	930	219	CNS00D23	
c 40	42.8	2.1	821	142	BE960968	601648521
c 41	42.8	2.1	1101	219	CNS0039L	
c 42	42.6	2.1	859	219	CNS004YY	
c 43	42.6	2.1	1077	222	CNS05AA9	
c 44	42.4	2.1	734	219	CNS010MP	
c 45	42.4	2.1	1092	220	CNS020K7	

ALIGNMENTS

RESULT 1
AQ319247/c
LOCUS
DEFINITION
DNA sequence.
ACCESSION
AQ319247
VERSION
AQ319247.1
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 654)
ADAM, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P., and Venter, J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL MEDLINE COMMENT Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-FT0175-310800-106-h09&t3=2000-08-31&t4=1)

Seq primer: puc 18 forward
High quality sequence stop: 290.

FEATURES source
1..290
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0175"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 78 a 73 c 53 g 86 t

ORIGIN

Query Match 12.8%; Score 260.4; DB 147; Length 290;
Best Local Similarity 97.8%; Pred. No. 6.4e-59;
Matches 264; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 740 gggaggagataaacacgggggcagagggtcaggattctggccctgctgctaaactgtgcg 799
Db 21 gggaggagataaacacgggggcagagggtcaggattctggccctgctgctaaactgtgcg 80
Qy 800 ttcataacccaatcatttcatttcttaacccctcaaaacaaagctgttgttaatactgat 859
Db 81 TTCATAACCAATCATTTTCATATTCTAACCCCTCAAAACAAAGCTGTGTAAATCTGAT 140
Qy 860 ctctacggttctcttggggcccaacattctccatataatccagccacacatcttttaata 919
Db 141 CTCTACGGTCTCTTCTGGGGCCCAACATCTCCATATATCCAGCCACACTCAATTTTAATA 200
Qy 920 ttttaattccagatctgactgtgacctttctacacttagaataacattactctatttg 979
Db 201 TTTAGTTCACAGATCTGACTGTGACCTTCTACACTGTAGAATAACATTTACTCAATTTGG 260
Qy 980 ttcaagacccttcgtgtgtgctgcctaata 1009
Db 261 TTCAAGACCCTTCGTGTGGCTAACGACTA 290

RESULT 8
BF373581/c 282 bp mRNA EST 24-NOV-2000
LOCUS BF373581 MR0-FT0175-210800-101-d05 FT0175 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF373581
ACCESSION BF373581
VERSION BF373581.1 GI:11335606
KEYWORDS EST.

SOURCE ORGANISM human. Homo sapiens

REFERENCE AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1. (bases 1 to 282)

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL MEDLINE COMMENT Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-FT0175-210800-101-d05&t3=2000-08-21&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 35
High quality sequence stop: 282.

FEATURES source
1..282
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0175"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 85 a 51 c 70 g 76 t

ORIGIN

Query Match 11.1%; Score 226.8; DB 147; Length 282;
Best Local Similarity 91.3%; Pred. No. 6.6e-50;
Matches 252; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

Qy 733 caagctggggaggagataaacacgggggcagagggtcaggattctgcccctgctgctaaa 792
Db 282 CAAGCTGGGAGGAGATAAACACGGGGCAGAGGGTTCAGGATTCTGGCCCTGCTGCTAAA 223
Qy 793 ctgtcgctcataacaaaatcatttcatttctaacccctcaaaacaaagctgttgttaat 852
Db 222 CTGTGCGTTCATAACCAATCATTTTCATATTCTAACCCCTCAAAACAAAGCTGTGTAAAT 163
Qy 853 atctgatctctacggttctcttctggtggcccaacattctccatataatccagccacactcatt 912
Db 162 ATCTGATCTCTACGGTCTCTTCTGGGGCCCAACATTTCTCATATATCCAGCCACACTCAAT 103
Qy 913 ttttaataatttagttccagatctgactgtgacctttc--tacactgtagaataacatta 970
Db 102 TTTAATATTTAGTTCACAGATCTGACTGTGACCTTTCCTACATCTGTGTAATAACATTA 43
Qy 971 ctccatttgttcaagacccttcgtgtgtgctgccta 1006
Db 42 CTCATTGTTCAAAGACCTTCACCTTTCACCTTGTAGCGACTA 7

RESULT 9

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 02:27:18 ; Search time 176.8 seconds
(without alignments)
2181.147 Million cell updates/sec

Title: US-09-402-713A-1
Perfect score: 2037
Sequence: 1 agaagtcgcatcagaaaa.....caataaagaatttacaaga 2037

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PT05_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59.8	2.9	7218	1 US-08-232-463-14	Sequence 14, Appl
2	57.4	2.8	7218	1 US-08-232-463-14	Sequence 14, Appl
3	47	2.3	10968	2 US-08-680-327-2	Sequence 2, Appl
4	47	2.3	10968	4 US-09-228-246-1	Sequence 1, Appl
5	45.4	2.2	5134	2 US-08-310-912A-157	Sequence 157, App
6	45.4	2.2	5134	5 PCT-US95-04589-157	Sequence 157, App
7	45.4	2.2	5475	2 US-08-680-327-1	Sequence 1, Appl
8	45.4	2.2	5475	4 US-09-228-246-3	Sequence 3, Appl
9	35.8	1.8	14507	3 US-08-785-150-1	Sequence 1, Appl
10	35	1.7	4129	2 US-08-370-319C-12	Sequence 12, Appl
11	35	1.7	4129	4 US-09-224-834-12	Sequence 12, Appl
12	34.8	1.7	5718	3 US-08-714-918-48	Sequence 48, Appl
13	34.8	1.7	5718	4 US-09-265-315-48	Sequence 48, Appl
14	34.8	1.7	5718	4 US-09-265-315-48	Sequence 48, Appl
15	34.8	1.7	5718	4 US-09-266-417-48	Sequence 48, Appl
16	34.2	1.7	2679	1 US-07-977-434-11	Sequence 11, Appl
17	34.2	1.7	2679	1 US-08-458-819-11	Sequence 11, Appl
18	34.2	1.7	2679	5 PCT-US91-07035-11	Sequence 11, Appl
19	33.8	1.7	2570	2 US-09-056-075-2	Sequence 20, Appl
20	33.8	1.7	246240	2 US-08-724-394A-20	Sequence 20, Appl
21	33.8	1.7	246240	2 US-08-724-394A-21	Sequence 21, Appl
22	33.8	1.7	246240	2 US-08-724-394A-22	Sequence 22, Appl
23	33.6	1.6	3084	1 US-08-551-437-3	Sequence 3, Appl
24	33.6	1.6	3084	3 US-09-004-225-3	Sequence 3, Appl
25	33.6	1.6	3084	3 US-09-084-346-3	Sequence 3, Appl
26	33.2	1.6	5173	1 US-08-242-677-1	Sequence 1, Appl
27	33	1.6	1952	1 US-08-333-358-1	Sequence 1, Appl

28	33	1.6	1952	1 US-08-463-694-1	Sequence 1, Appl
29	33	1.6	1952	1 US-08-694-501-1	Sequence 1, Appl
30	33	1.6	2950	5 PCT-US93-08386-7	Sequence 7, Appl
31	33	1.6	2968	5 PCT-US93-08386-1	Sequence 1, Appl
32	33	1.6	8920	2 US-08-446-855A-1	Sequence 1, Appl
33	33	1.6	8920	4 US-09-150-741-1	Sequence 1, Appl
34	33	1.6	720	4 US-08-446-935-4	Sequence 4, Appl
35	32.8	1.6	2395	4 US-08-446-935-7	Sequence 7, Appl
36	32.8	1.6	2405	1 US-08-454-097-30	Sequence 30, Appl
37	32.8	1.6	2405	3 US-08-185-359-30	Sequence 30, Appl
38	32.8	1.6	5356	4 US-08-446-935-1	Sequence 1, Appl
39	32.6	1.6	1400	1 US-07-989-845-1	Sequence 1, Appl
40	32.6	1.6	1400	5 PCT-US93-11298-1	Sequence 1, Appl
41	32.6	1.6	3607	1 US-08-647-351B-1	Sequence 1, Appl
42	32.6	1.6	246240	2 US-08-724-394A-20	Sequence 20, Appl
43	32.6	1.6	246240	2 US-08-724-394A-21	Sequence 21, Appl
44	32.6	1.6	246240	2 US-08-724-394A-22	Sequence 22, Appl
45	32	1.6	713	3 US-08-532-896-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
; US-08-232-463-14

```
Query Match 2.9%; Score 59.8; DB 1; Length 7218;
Best Local Similarity 5.1%; Pred. No. 1.2e-07;
Matches 20; Conservative 219; Mismatches 155; Indels 0; Gaps 0;

QY 1178 ttcattccacatttggcccaattctcaagacacctcaaaatgcattccattcaataatcac 1237
Db 1076 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1135
QY 1238 aggattaaactttttttaacctggaagaatcaatgttacatgcagctatgggaattt 1297
Db 1136 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1195
QY 1298 aattacataatttggtttcocagtcgaagatgactaagtccttattacocctccttggtt 1357
Db 1196 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1255
QY 1358 tgattttttccagataaagttaaaatgcttagcctgtactgtgactgagctgtatagca 1417
Db 1256 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1315
QY 1418 cagcctctcccatccctccagccttatctgtcatcaccatcaacccctcccatnysacc 1477
Db 1316 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1375
QY 1478 taacaaaatctaacttgaattccttgtaacatgtcaggnacatacttctctgc 1537
Db 1376 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1435
QY 1538 tgagaagctctctctgtctcttaanttagaat 1571
Db 1436 TACCAAAATCTTCTATCTCTTTAACTACTTGCAT 1469
```

```
RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgt-Fls
US-08-232-463-14

Query Match 2.8%; Score 57.4; DB 1; Length 7218;
Best Local Similarity 4.3%; Pred. No. 6.4e-07;
Matches 16; Conservative 211; Mismatches 142; Indels 0; Gaps 0;

QY 422 aggaagacacaaaggaagcacagagatccctgggagaaatgccggccgccatcttgggt 481
Db 1407 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1348
QY 482 catgatgagctgcctctgctgctgctgctgctgctgctgctgctgctgctgctgctg 541
Db 1347 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1288
QY 542 attgatgttctctaaagatggcaggaacagatcctgtgtgatatatttga 601
Db 1287 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1228
QY 602 acgggattcacagattgaaatgaagtcaacaagtgcattaccatgagagaaacag 661
Db 1227 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1168
QY 662 acgagaaaatcttgatggcttcacaagacatgcacaaacaaatggaatactgtgatga 721
Db 1167 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1108
QY 722 catgagcgagcgaagctggggaggagataaacacggggcagagggcagagattctggccc 781
Db 1107 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1048
QY 782 tgctgccta 790
Db 1047 TGCAGCCAA 1039

RESULT 3
US-08-680-327-2/c
; Sequence 2, Application US/08680327
; Patent No. 5859321
; GENERAL INFORMATION:
; APPLICANT: Staskawicz, Brian S., Oldroyd, Giles Edward,
; APPLICANT: Salmeron, John M., Rommens, Caius
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT
; TITLE OF INVENTION: PATHOGEN RESISTANCE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist
; ADDRESSEE: Whinston
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,327
; FILING DATE: July 11, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
```


APPLICANT: MINDILINGS, MICHAEL N.
 APPLICANT: YU, GUO-LIANG
 TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND DETECTION
 TITLE OF INVENTION: METHODS
 NUMBER OF SEQUENCES: 208
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2904
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/310,912A
 FILING DATE: September 22, 1994
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/227,360

;; FILING DATE:;;
;; CLASSIFICATION:;;
;; PRIOR APPLICATION DATA:;;
;; APPLICATION NUMBER: US/08/586,509;;
;; FILING DATE: 11-JAN-96;;
;; ATTORNEY/AGENT INFORMATION:;;
;; NAME: Perkins, Patricia Anne;;
;; REGISTRATION NUMBER: 34,693;;
;; REFERENCE/DOCKET NUMBER: 2841;;
;; TELECOMMUNICATION INFORMATION:;;
;; TELEPHONE: (206)587-0430;;
;; TELEFAX: (206)233-0644;;
;; INFORMATION FOR SEQ ID NO: 1:;;
;; SEQUENCE CHARACTERISTICS:;;
;; LENGTH: 14507 base pairs;;
;; TYPE: nucleic acid;;
;; STRANDEDNESS: single;;
;; TOPOLOGY: NO. 6027915 Relevant;;
;; MOLECULE TYPE: DNA (genomic);;
;; HYPOTHETICAL: NO;;
;; ANTI-SENSE: NO;;
;; ORIGINAL SOURCE:;;
;; ORGANISM: Chinese hamster;;
;; IMMEDIATE SOURCE:;;
;; CLONE: 2A5-3 lambda CHO sequence;;
US-08-785-150-1

Query Match 1.8%; Score 35.8; DB 3; Length 14507;
Best Local Similarity 46.9%; Pred. No. 3.1; Indels 0; Gaps 0;
Matches 112; Conservative 0; Mismatches 127;
QY 1035 gtgttctggccagggtatctgtgaacaggctgtgggaagcatctcaagatctttccagggt 1094
Db 10959 GGGTCTTCCACAGTGATGATCTCATCAAGGGAATTTATTCACGCCCTGGTGTAGTG 10900
QY 1095 tatactactagcacagcatgatcattacaggagtggaattatcttaatacaacatcatcct 1154
Db 10899 GCTACAACTAGCCGGCTAGAGGCGACGACGCCCTTCTTAATCAGGCTCTGTGTGTAC 10840
QY 1155 cagtgtcttgcccatctgaattcatcttccacatttggccatttgcctcctcaagacctca 1214
Db 10839 GACGATGACTAACGCCAGTAGTAACAGGGGAACCTGTTGAGAATTTAAAGTGCTCA 10780
QY 1215 aaatgcattcccaatcaatcagcagattaaacttttttttttaacctggaagaattcaa 1273
Db 10779 CAACGTGATTTCAGTAGTAGACAGCAGAAACATTTATTTTATTTGGAAATAATATA 10721

RESULT 10
US-08-370-319C-12/C
; Sequence 10, Application US/08370319C
; Patent No. 5856091
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; APPLICANT: Traversari, Catia; W lfel, Thomas; Coullie, Pierre;
; APPLICANT: Boon-Falleur, Thierry; De Plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE TU
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect

;; CURRENT APPLICATION DATA:;;
;; APPLICATION NUMBER: US/08/370,319C;;
;; FILING DATE: 10-JANUARY-1995;;
;; CLASSIFICATION: 435;;
;; PRIOR APPLICATION DATA:;;
;; APPLICATION NUMBER: 08/272,351;;
;; FILING DATE: 8-JULY-1994;;
;; PRIOR APPLICATION DATA:;;
;; APPLICATION NUMBER: 08/032,978;;
;; FILING DATE: 18-MAR-1993;;
;; ATTORNEY/AGENT INFORMATION:;;
;; NAME: Hanson, No. 5856091man D.;;
;; REGISTRATION NUMBER: 30,946;;
;; REFERENCE/DOCKET NUMBER: LUD 5377.1;;
;; TELECOMMUNICATION INFORMATION:;;
;; TELEPHONE: (212) 688-9200;;
;; TELEFAX: (212) 838-3884;;
;; INFORMATION FOR SEQ ID NO: 12:;;
;; SEQUENCE CHARACTERISTICS:;;
;; LENGTH: 4129 base pairs;;
;; TYPE: nucleic acid;;
;; STRANDEDNESS: double;;
;; TOPOLOGY: linear;;
;; FEATURE:
;; OTHER INFORMATION: The sequence is preceded by an
;; OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3
;; OTHER INFORMATION: kilobases
US-08-370-319C-12

Query Match 1.7%; Score 35; DB 2; Length 4129;
Best Local Similarity 48.7%; Pred. No. 2.5;
Matches 95; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
QY 1190 ttttggccattctcaagacctcaaaatgtcattccattcaataatcacaggattaaacttt 1249
Db 2703 TCTTTCTCATACTAATAATTTTTTGAATATATGATATGATATTAATATACATGCTTATTTCA 2644
QY 1250 ttttttaacctgggaagaattcaattgtacatgcagctatgggaatttaattacatttt 1309
Db 2643 AACTCCTGTGTTTAACTGTTCAATAGCACCATGTGCTAGGCTAAGCCATCAGACAG 2584
QY 1310 tttttccagtgcgaagatgactaagtccttccctccctcccttgggtgatttttttc 1369
Db 2583 CACAGCTCTATACATAAATAATAGTACTTATCCCTCCCTCTGTATATATATCTGCATC 2524
QY 1370 cagtataaagttaa 1384
Db 2523 CATTAAAGGCAAA 2509

RESULT 11
US-09-224-834-12/C
; Sequence 12, Application US/09224834
; Patent No. 6201111
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; APPLICANT: Traversari, Catia; W lfel, Thomas; Coullie, Pierre;
; APPLICANT: Boon-Falleur, Thierry; De Plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; OPERATING SYSTEM: IBM PS/2
; SOFTWARE: Wordperfect

APPLICANT: Martin, Patrick K.
 APPLICANT: Schmid, Molly B.
 APPLICANT: Sun, Dongxu
 TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
 TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
 TITLE OF INVENTION: TARGET GENES
 NUMBER OF SEQUENCES: 111
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Lyon & Lyon
 STREET: 633 West Fifth Street
 STREET: Suite 4700
 CITY: Los Angeles
 STATE: California
 COUNTRY: U.S.A.
 ZIP: 90071-2066
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 MEDIUM TYPE: storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: IBM P.C. DOS 5.0
 SOFTWARE: Word Perfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/266,417
 FILING DATE: March 9, 1999
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/714,918
 FILING DATE: September 13, 1996
 APPLICATION NUMBER: 60/009,102
 FILING DATE: December 22, 1995
 APPLICATION NUMBER: 60/003,798
 FILING DATE: September 15, 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Warburg, Richard J.
 REGISTRATION NUMBER: 32,327
 REFERENCE/DOCKET NUMBER: 240/248
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (213) 489-1600
 TELEFAX: (213) 955-0440
 TELEX: 67-3510
 INFORMATION FOR SEQ ID NO: 48:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5718 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-266-417-48

Query Match 1.7%; Score 34.8; DB 4; Length 5718;
 Best Local Similarity 48.1%; Pred. No. 3.5;
 Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
 QY 1198 ccattctcaagaccctcaaaatgcattccatttaataatcacaggattacatttttttta 1257
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 DB 283 CCATCAGTAATAACGACACTGATTTCCTTAATTGTGTAATCATATACATTTTCTTTTA 342
 QY 1258 acctggaagaattcaatgttacatgcagctatggaaatttaattacatatattgtttcc 1317
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 DB 343 TCTTCATAAATATCTTACACGGTTCAAGCAGCCAGGTGATGCTAAACTTAATTC 402
 QY 1318 agtgcagaagatgaactaaagtcctttatccctccctttgtttgattttttccagtataa 1377
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 DB 403 TCTTTATTAGTAACITTTACATTTGGTTTAACTTCTAATTACCTTGATTACGTTTGTC 462
 QY 1378 agttaaaatgcttagcctgtactga 1403
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 DB 463 ATTTCGAATGCTTCATCTCTTAATGA 488

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 02:29:48 ; Search time 478.87 Seconds
(without alignments)
2670.945 Million cell updates/sec

Title: US-09-402-713a-1

Perfect score: 2037

Sequence: 1 agaagctggcatcagaaaaa.....caataaagaatttacaaga 2037

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2032	99.8	2037	19	AAV62427
2	2030.4	99.7	3582	19	AAV62430
3	1772.4	87.0	1872	19	AAV62428
4	1737.8	85.3	3112	21	AAA06687
5	1731.4	85.0	2426	21	AAA06689
6	1712.4	84.1	2329	21	AAA06688
7	725.4	35.6	820	19	AAV62429
8	722.4	35.5	812	21	AAA06690
9	524	25.7	597	20	AAV37486
10	457.2	22.4	718	21	AAA06545
11	319.6	15.7	437	21	AAV06768

c	12	288.4	14.2	301	21	AAA06520	Human immunogenic
c	13	235.4	11.6	359	20	AAZ33445	Human prostate can
c	14	136	6.7	936	22	AAF58252	Oligonucleotide D1
c	15	136	6.7	936	22	AAF58254	Oligonucleotide D1
c	16	136	6.7	936	22	AAF58257	Oligonucleotide D1
c	17	136	6.7	936	22	AAF58259	Oligonucleotide D2
c	18	136	6.7	936	22	AAF58262	Oligonucleotide D2
c	19	136	6.7	936	22	AAF58255	Oligonucleotide D1
c	20	134.4	6.6	936	22	AAF58252	Oligonucleotide D1
c	21	134.4	6.6	936	22	AAF58254	Oligonucleotide D1
c	22	134.4	6.6	936	22	AAF58257	Oligonucleotide D1
c	23	134.4	6.6	936	22	AAF58259	Oligonucleotide D2
c	24	134.4	6.6	936	22	AAF58262	Oligonucleotide D2
c	25	134.4	6.6	936	22	AAF58255	Oligonucleotide D1
c	26	99	4.9	123	19	AAV3791	Prostate cancer an
c	27	50.8	2.5	244	22	AAF58238	Oligonucleotide D1
c	28	47.6	2.3	244	22	AAF58238	Oligonucleotide D1
c	29	47	2.3	10968	19	AAV17789	Tomato Prf genomic
c	30	45.4	2.2	5475	19	AAV17777	Tomato Prf cDNA.
c	31	40	2.0	796	20	AAV37484	Human secreted pro
c	32	39	1.9	2206	21	AAV12390	Human interferon-b
c	33	39	1.9	10325	20	AAV76516	Plasmid containing
c	34	37.6	1.8	1297	13	AAQ21645	3' coding sequence
c	35	37.6	1.8	1770	14	AAQ37684	P.falciparum antig
c	36	37.6	1.8	1770	14	AAQ37124	P.falciparum antig
c	37	37.4	1.8	1255	22	AAF72762	Human prostate can
c	38	37.4	1.8	1325	21	AAF15730	Human prostate can
c	39	37.4	1.8	1325	22	AAF72787	Human prostate can
c	40	37.2	1.8	1751	21	AAC68122	Human secreted pro
c	41	36.8	1.8	5804	18	AAV74560	Staphylococcus aur
c	42	36.4	1.8	2643	15	AAQ66096	S. cerevisiae Plc1
c	43	36.2	1.8	1451	21	AAF21739	Human breast and o
c	44	36.2	1.8	1517	21	AAC34792	Arabidopsis thalia
c	45	36.2	1.8	1630	20	AAV13310	Enterococcus faeca

ALIGNMENTS

RESULT 1

AAV62427
ID AAV62427 standard; cDNA; 2037 BP.

XX
AC AAV62427;

XX
DT 30-DEC-1998 (first entry)

DE
XX Prostate cancer antigen (PCA3) cDNA splice variant 1.

XX
KW Prostate cancer antigen cDNA splice variant 1; PCA3; prostatic cancer;
PC; ds.

XX
OS Homo sapiens.

XX
FH Key Location/Qualifiers
FT CDS 379..534

FT
FT /tag= a
/product= "PCA3 protein variant 1"

FT
FT polyA_signal 2019..2024
/tag= b

XX
PN WO9845420-A1.

XX
PD 15-OCT-1998.

XX
PF 09-APR-1998; 98WO-CA00346.

XX
PR 10-APR-1997; 97US-0041836.

XX
PA (DIAG-) DIAGNOCURE INC.

XX
PI Bussemakers MJG;

DR WPI; 1998-568347/48.
XX P-PSDB; AAW79736.
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
XX
PS
PS Claim 3; Fig 2B-2J; 11lpp; English.
XX
CC The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 1 sequence comprising of exons 1', 2, 3, 4a and
CC 4b of the PCA3 gene. The PCA3 cDNA splice variant 1 sequence,
CC isolated from a human primary prostatic tumour tissue cDNA library,
CC was found in approximately 5% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunoassay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (ant)agonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
XX Sequence 2037 BP; 622 A; 426 C; 406 G; 575 T; 8 other;

Query Match 99.8%; Score 2032; DB 19; Length 2037;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2037; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaagctggcatcagaaaaaacagagggagattgtgtgctgcagcccgagggagaccag 60
DB 1 ageagctggcatcagaaaaaacagagggagattgtgtgctgcagcccgagggagaccag 60
QY 61 gaagatctcattgtgggaagacctgatgatacagaggaattacaacacatactatg 120
DB 61 gaagatctcattgtgggaagacctgatgatacagaggaattacaacacatactatg 120
QY 121 tgtttcaatgaacacacagataaataagtgaaagagctagtcgcgtgagctccctcagt 180
DB 121 tgtttcaatgaacacacagataaataagtgaaagagctagtcgcgtgagctccctcagt 180
QY 181 gacacagggctggatcaccatcgacggcactttctgagtactcagtgagcaagaaga 240
DB 181 gacacagggctggatcaccatcgacggcactttctgagtactcagtgagcaagaaga 240
QY 241 ctacagacatcctcaatggcaggggtgagaaataagaaggctgctgactttaccatctga 300
DB 241 ctacagacatcctcaatggcaggggtgagaaataagaaggctgctgactttaccatctga 300
QY 301 ggcacacatctgctgaatggagataataacatacactagaaacagcaagatgacaata 360
DB 301 ggcacacatctgctgaatggagataataacatacactagaaacagcaagatgacaata 360
QY 361 taatgtctaagtagtagatgtttttgacattttccagcccttttaatatccacacaca 420
DB 361 taatgtctaagtagtagatgtttttgacattttccagcccttttaatatccacacaca 420
QY 421 caggaagcacaagaagaaacagagatccctgggagaaatcccgcccgccatcttggg 480
DB 421 caggaagcacaagaagaaacagagatccctgggagaaatcccgcccgccatcttggg 480
QY 481 tcatgatgagctgcctcctgtgctcctgtgctcctgtgaggaagacattagaaaaatg 540
DB 481 tcatgatgagctgcctcctgtgctcctgtgctcctgtgaggaagacattagaaaaatg 540
QY 541 aattgatgtgttcttaaggatgggcaggaagaaacagatcctgtgtgatatatttg 600

DB 541 aattgatgtgttcttaaggatgggcaggaagaaacagatcctgtgtgatatatttg 600
QY 601 aacgggattacagatttgaatgaagtcacaagtgagcattaccatgagaggaataa 660
DB 601 aacgggattacagatttgaatgaagtcacaagtgagcattaccatgagaggaataa 660
QY 661 gacgagaaatcttgatggcttcacaagacatgcaacaaacaaatggaaatctgtatg 720
DB 661 gacgagaaatcttgatggcttcacaagacatgcaacaaacaaatggaaatctgtatg 720
QY 721 acatgaggcagccaagctgggaggagataaacacaggggcagaggtcaggattctggcc 780
DB 721 acatgaggcagccaagctgggaggagataaacacaggggcagaggtcaggattctggcc 780
QY 781 ctgctgcctcaactgtgccttcataacacaaatcatttcatttccaccccaaaaaca 840
DB 781 ctgctgcctcaactgtgccttcataacacaaatcatttcatttccaccccaaaaaca 840
QY 841 agctgtgtaatatctgatctctacaggttccctctggtggcccaacattctccatatcca 900
DB 841 agctgtgtaatatctgatctctacaggttccctctggtggcccaacattctccatatcca 900
QY 901 gccacactcatttttaataatttagttccagatctgtactgtgacctttctacactgtg 960
DB 901 gccacactcatttttaataatttagttccagatctgtactgtgacctttctacactgtg 960
QY 961 aataacattactcattttgttcaaaagacctctgtgttgccttaataatgactgact 1020
DB 961 aataacattactcattttgttcaaaagacctctgtgttgccttaataatgactgact 1020
QY 1021 gtttttccctaaaggagtgtctggtggccaggggactctggaacaggtgggaagcatctcaa 1080
DB 1021 gtttttccctaaaggagtgtctggtggccaggggactctggaacaggtgggaagcatctcaa 1080
QY 1081 gatctttccaggggttatacttacttagcacacagcatgatactacggagtgaaattatcta 1140
DB 1081 gatctttccaggggttatacttacttagcacacagcatgatactacggagtgaaattatcta 1140
QY 1141 atcaacatcatcctcagtgctttgtggccatactgaaatcatttccacattttgtgccca 1200
DB 1141 atcaacatcatcctcagtgctttgtggccatactgaaatcatttccacattttgtgccca 1200
QY 1201 ttctcaagacctcacaatgtctatccattcaatcacaggtataacttttttttaacc 1260
DB 1201 ttctcaagacctcacaatgtctatccattcaatcacaggtataacttttttttaacc 1260
QY 1261 tggagaattcaatgttatacagcagctatgggaatttaattacataattttgttttcag 1320
DB 1261 tggagaattcaatgttatacagcagctatgggaatttaattacataattttgttttcag 1320
QY 1321 gcaagagatgaactaagtcctttatccctccctctgtttgtttttttccagataaagt 1380
DB 1321 gcaagagatgaactaagtcctttatccctccctctgtttgtttttttccagataaagt 1380
QY 1381 taaaatgcttagcctttagctagggctgtatcacagcacagcctctcccatccctccag 1440
DB 1381 taaaatgcttagcctttagctagggctgtatcacagcacagcctctcccatccctccag 1440
QY 1441 ctatatgtcatcacaatcacaacccctcccatnysacctaaacaaatcacttaacttgaatt 1500
DB 1441 ctatatgtcatcacaatcacaacccctcccatnysacctaaacaaatcacttaacttgaatt 1500
QY 1501 ccttgaacatgcagcagncatatactttctctgctgagagctcttccctgtctctt 1560
DB 1501 ccttgaacatgcagcagncatatactttctctgctgagagctcttccctgtctctt 1560
QY 1561 aantctagaatgatgataaagttttgaataagtgactatcttacttcatgcgaaagaagg 1620
DB 1561 aantctagaatgatgataaagttttgaataagtgactatcttacttcatgcgaaagaagg 1620
QY 1621 acacatatgagattcatcatcatcatgagacgaaataactaaagtgtaattgtattata 1680
DB 1621 acacatatgagattcatcatcatcatgagacgaaataactaaagtgtaattgtattata 1680

Db 1621 acacatatgagattcatcatcacatgagacagcaataactactaaaagttaattgattata 1680
Qy 1681 agagtttagataaataatgaatgcaagkaccacagaggaatgtttatgggcacgtt 1740
Db 1681 agagtttagataaataatgaatgcaagkaccacagaggaatgtttatgggcacgtt 1740
Qy 1741 tgaagcctggatgtgaagmaaaggcagggaacctcatagtatcttataataatactt 1800
Db 1741 tgaagcctggatgtgaagmaaaggcagggaacctcatagtatcttataataatactt 1800
Qy 1801 catttctctctctatcatcaaatcaacagctttttcacagaattcatcagtgcaaa 1860
Db 1801 catttctctctctatcatcaaatcaacagctttttcacagaattcatcagtgcaaa 1860
Qy 1861 tccccaaaggtaacctttatcattcatgtgagtcgcttttagaattttggcaaatca 1920
Db 1861 tccccaaaggtaacctttatcatttcatgtgagtcgcttttagaattttggcaaatca 1920
Qy 1921 tactggtcacttatctcaactttgagatgtgtttgtcctttagttaattgaaagaaata 1980
Db 1921 tactggtcacttatctcaactttgagatgtgtttgtcctttagttaattgaaagaaata 1980
Qy 1981 gggcactcttgagcacttttagggttcactcctctggaataaagaatttacaaga 2037
Db 1981 gggcactcttgagcacttttagggttcactcctctggaataaagaatttacaaga 2037

RESULT 2

ID AAV62430 standard; cDNA; 3582 BP.

AC AAV62430;

30-DEC-1998 (first entry)

Prostate cancer antigen (PCA3) wild-type cDNA.

Prostate cancer antigen cDNA; PCA3; prostatic cancer;

PC; ds.

Homo sapiens.

Key Location/Qualifiers

CDS 401..556

/*tag= a

/product= "PCA3 protein"

983..987

polyA_signal

/*tag= b

2041..2046

polyA_signal

/*tag= c

2597..2602

polyA_signal

/*tag= d

3494..3496

polyA_signal

/*tag= e

W09845420-A1.

15-OCT-1998.

PS Claim 3; Fig 5B-5F; 11pp; English.
XX The present sequence represents the prostate cancer antigen (PCA3)
CC wild-type cDNA sequence comprising of exons 1, 2, 3, 4a-4d of the
CC PCA3 gene. The invention claims for PCA3 cDNA variants and the
CC proteins they encode. The invention also claims for antibodies
CC against PCA3 protein. The antibodies are claimed to be useful for
CC detecting PCA3 protein in immunoassay tests, for diagnosing, assessing
CC and prognosing of prostatic cancer (PC). Antibodies, optionally
CC coupled to a cytotoxin or radioisotope, and nucleic acids antisense
CC to PCA3 cDNA are claimed to be useful for treating PC, while determining
CC elevated levels of PCA3 (as RNA or protein) is useful for detecting a
CC predisposition to development of PC, e.g. in prenatal tests. Detecting
CC PCA3 protein allows differentiation between malignant and benign
CC prostatic disease, and the level of PCA3 expression allows correlation
CC with the grade of tumour. PCA3 protein and its fragments are also
CC claimed to be useful in vaccines for preventing PC; in drug screens
CC for identifying specific (ant)agonists (potentially useful
CC therapeutically) and for studying protein-DNA interactions.
XX
SQ Sequence 3582 BP; 1052 A; 788 C; 679 G; 1063 T; 0 other;

Query Match 99.7%; Score 2030.4; DB 19; Length 3582;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2028; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 agaagctggcatcagaaaaacacagagggagatttgtgtggctgacgcagggagaccag 60
Db 23 agaagctggcatcagaaaaacacagagggagatttgtgtggctgacgcagggagaccag 82
Qy 61 gaagatctgcattgtgggaagacctgatgacagaggaattacaacacatacttag 120
Db 83 gaagatctgcattgtgggaagacctgatgacagaggaattacaacacatacttag 142
Qy 121 tgttcaatgaacacccaagataaaataagtagagtagtcgcgtgtgagtcctcagt 180
Db 143 tgttcaatgaacacccaagataaaataagtagagtagtcgcgtgtgagtcctcagt 202
Qy 181 gacacagggctggatccaccatcgacgcgcactttctgactactcagcagaaagaga 240
Db 203 gacacagggctggatccaccatcgacgcgcactttctgactactcagcagaaagaga 262
Qy 241 ctacagacatctcaatggcaggggtgagaaataaagaaggctgctgactttaccatctga 300
Db 263 ctacagacatctcaatggcaggggtgagaaataaagaaggctgctgactttaccatctga 322
Qy 301 ggcacacatctgctgaaatggagataataacatcactagaacacagcaagatgacaata 360
Db 323 ggcacacatctgctgaaatggagataataacatcactagaacacagcaagatgacaata 382
Qy 361 taatgtctaaagttagatgtttttgacattttccagccctttaaatacacacaca 420
Db 383 taatgtctaaagttagatgtttttgacattttccagccctttaaatacacacaca 442
Qy 421 caggaagcacaaaaaggaagcagagatccctgggagaaatgcccgcccatcttggg 480
Db 443 caggaagcacaaaaaggaagcagagatccctgggagaaatgcccgcccatcttggg 502
Qy 481 tcatcgatgagcctcgccctgtgctgtgctcccgctttgaggggaaggacattagaaaaatg 540
Db 503 tcatcgatgagcctcgccctgtgctgtgctcccgctttgaggggaaggacattagaaaaatg 562
Qy 541 aattgatgtgttccttaaaggatgggcagggaacacagatcctgttggatattttttg 600
Db 563 aattgatgtgttccttaaaggatgggcagggaacacagatcctgttggatattttttg 622
Qy 601 aacgggattcacagattgaaatgaagtacaaagtgacattaccatgagaggaagaaaca 660
Db 623 aacgggattcacagattgaaatgaagtacaaagtgacattaccatgagaggaagaaaca 682
Qy 661 gacgagaaaaatcttgatgcttcacagacatgcaacaaacaaaatggaaactactgtg 720

New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
prevention and treatment of prostatic cancer

Db 683 gacgagaaaatcttgatggttcacaaagacatgcaacaaacaaaatggaatactgtgatg 742
QY 721 acatgagcagcgaagctgggagagagataaacacagggcgagaggtcaggattctggcc 780
Db 743 acatgagcagcgaagctgggagagagataaacacagggcgagaggtcaggattctggcc 802
QY 781 ctgctgctaaactgctgcttataacaaaatcatttctatattttaaaccctcaaaacaa 840
Db 803 ctgctgctaaactgctgcttataacaaaatcatttctatattttaaaccctcaaaacaa 862
QY 841 agctgtgtaatactgatctctacaggttctctctgggcccacattctccatatccaa 900
Db 863 agctgtgtaatactgatctctacaggttctctctgggcccacattctccatatccaa 922
QY 901 gccacactatattttaaattttagtctccagatctgactgtgaaccttctcaactgtag 960
Db 923 gccacactatattttaaattttagtctccagatctgactgtgaaccttctcaactgtag 982
QY 961 aataacattactactttgttcaaaagacccttcgtgtgctgcctaatatgtagctgact 1020
Db 983 aataacattactactttgttcaaaagacccttcgtgtgctgcctaatatgtagctgact 1042
QY 1021 gtttttccaaagagtgcttggccaggggagctctgtgaacaggtgggaagcattctcaa 1080
Db 1043 gtttttccaaagagtgcttggccaggggagctctgtgaacaggtgggaagcattctcaa 1102
QY 1081 gatcttccaggggttactacttagcacacagcatgatacagagtggaattatctta 1140
Db 1103 gatcttccaggggttactacttagcacacagcatgatacagagtggaattatctta 1162
QY 1141 atcaacatcatctcagtgcttggccatactgaaattcatttcccaactttgtgtccca 1200
Db 1163 atcaacatcatctcagtgcttggccatactgaaattcatttcccaactttgtgtccca 1222
QY 1201 ttctcaagacctcaaaatgctcattcatttaataacacagattaaactttttttaacc 1260
Db 1223 ttctcaagacctcaaaatgctcattcatttaataacacagattaaactttttttaacc 1282
QY 1261 tggagaattcaatgttacatgcagctatgggaatttaattacataatttgtttccagt 1320
Db 1283 tggagaattcaatgttacatgcagctatgggaatttaattacataatttgtttccagt 1342
QY 1321 gcaagatgactaaagtcctttatccctccctcttgtgtgattttttccagataaaagt 1380
Db 1343 gcaagatgactaaagtcctttatccctccctcttgtgtgattttttccagataaaagt 1402
QY 1381 taaaatgcttagcctgtactaggtgtatagcagacagcctctcccatccctccagc 1440
Db 1403 taaaatgcttagcctgtactaggtgtatagcagacagcctctcccatccctccagc 1462
QY 1441 ctatctgtcatcaccatcaacccctcccatnysacctaaacaaatctaaacttgaatt 1500
Db 1463 ctatctgtcatcaccatcaacccctcccatcaccacaaacaaatctaaacttgaatt 1522
QY 1501 ccttgaacatgctcaggnacatattttctctctgctgagagctcttctctgtctctt 1560
Db 1523 ccttgaacatgctcaggaacatatttctctctgctgagagctcttctctgtctctt 1582
QY 1561 aantctagaatgatgaatttgaattgaattgaattgactattctactcatcgaagaagg 1620
Db 1583 aantctagaatgatgaatttgaattgaattgaattgactattctactcatcgaagaagg 1642
QY 1621 acacatagagattcatcatcacatgagacagcaaaatactaaagtgaatttgattata 1680
Db 1643 acacatagagattcatcatcacatgagacagcaaaatactaaagtgaatttgattata 1702
QY 1681 agagtttagataataatgaatgaagagkccacagagggaatttttatggggcacgtt 1740
Db 1703 agagtttagataataatgaatgaagagccacagagggaatttttatggggcacgtt 1762
QY 1741 tgaagctggagtggaamaagcgagggaacctatagttatttataataactt 1800
Db 1763 tgaagctggagtggaagcaaaagcgagggaacctatagttatttataataactt 1822

QY 1801 catttctctatctctatccacaatatccacaagctttttcacagaattcatgctgcaaa 1860
Db 1823 catttctctatctctatccacaatatccacaagctttttcacagaattcatgctgcaaa 1882
QY 1861 tccccaaagtaaacctttatccatttcatgtgagtgcgcttttagaattttggcaaatca 1920
Db 1883 tccccaaagtaaacctttatccatttcatgtgagtgcgcttttagaattttggcaaatca 1942
QY 1921 tactgtcacttatctcacaactttgagatgtgttgccttctgttagttaaagaagaata 1980
Db 1943 tactgtcacttatctcacaactttgagatgtgttgccttctgttagttaaagaagaata 2002
QY 1981 gggcactcttgagcacttttaggttcaactctgccaataaagaattttacaaga 2037
Db 2003 gggcactcttgagcacttttaggttcaactctgccaataaagaattttacaaga 2059
RESULT 3
AAV62428
ID AAV62428 standard; cDNA; 1872 BP.
XX
AC AAV62428;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 2.
XX
KW Prostate cancer antigen cDNA splice variant 2; PCA3; prostatic cancer;
PC; ds.
XX
OS Homo sapiens.
XX
PN W09845420-AL.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
prevention and treatment of prostatic cancer
PS Claim 4; Pages 76-77; 11lpp; English.
XX
CC The present sequence represents the prostate cancer antigen (PCA3)
cDNA splice variant 2 sequence comprising of exons 1, 3, 4a and
4b of the PCA3 gene. The PCA3 cDNA splice variant 2 sequence,
isolated from a human primary prostatic tumour tissue cDNA library,
was found in approximately 65% of the cDNA clones isolated. The
invention claims for PCA3 cDNA variants and the proteins they encode.
The invention also claims for antibodies against PCA3 protein. The
antibodies are claimed to be useful for detecting PCA3 protein in
immunossay tests, for diagnosing, assessing and prognosing of
prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
to be useful for treating PC, while determining elevated levels of
PCA3 (as RNA or protein) is useful for detecting a predisposition
to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
allows differentiation between malignant and benign prostatic disease,
and the level of PCA3 expression allows correlation with the grade of
tumour. PCA3 protein and its fragments are also claimed to be useful
in vaccines for preventing PC; in drug screens for identifying
specific (antagonists (potentially useful therapeutically) and for
studying protein-DNA interactions.
XX

SQ Sequence 1872 BP; 567 A; 389 C; 369 G; 539 T; 8 other;

Query Match	87.08	Score 1772.4	DB 19	Length 1872
Best Local Similarity	99.99	Pred. No. 0		
Matches 1778	Conservative	0	Mismatches	1
			Indels	0
			Gaps	0
Qy	259	caggggtgagaaaTaagaaagcgtgctgactttaccatctgagggccacacatctgctgaa	318	
Db	94	cagaggtgagaaaTaagaaagcgtgactttaccatctgagggccacacatctgctgaa	153	
Qy	319	atggagataattaacatcactagaaacagcaagatgacaataataatgtctaaagtgtgac	378	
Db	154	atggagataattaacatcactagaaacagcaagatgacaataataatgtctaaagtgtgac	213	
Qy	379	atgtttttgacatttccagccctttaaatatccacacacaggaagcacaagaagaa	438	
Db	214	atgtttttgacatttccagccctttaaatatccacacacaggaagcacaagaagaa	273	
Qy	439	gcacagagatcccttgaggaaatgcccggccgcacatcttggtcatgatgagcctcgcc	498	
Db	274	gcacagagatcccttgaggaaatgcccggccgcacatcttggtcatgatgagcctcgcc	333	
Qy	499	ctgtgcctgttcccgcttgtgaggaaagacattagaaaaatgaattgatgtgttccttaa	558	
Db	334	ctgtgcctgttcccgcttgtgaggaaagacattagaaaaatgaattgatgtgttccttaa	393	
Qy	559	aggatggcagaaaaacagatcctgtgtgatatatttttgaacgggattacagatttg	618	
Db	394	aggatggcagaaaaacagatcctgtgtgatatatttttgaacgggattacagatttg	453	
Qy	619	aaatgaagtcaaaagttagcatcaccatgagagaaacagacgagaaatcttgatg	678	
Db	454	aaatgaagtcaaaagttagcatcaccatgagagaaacagacgagaaatcttgatg	513	
Qy	679	gcttcacaagacatgcaacaaacaaaaatggaaatctgtgtgacatgaggcagccaagt	738	
Db	514	gcttcacaagacatgcaacaaacaaaaatggaaatctgtgtgacatgaggcagccaagt	573	
Qy	739	ggggaggagataaacacaggggcagaggctcaggattctggccctgctgcctaaactgtgc	798	
Db	574	ggggaggagataaacacaggggcagaggctcaggattctggccctgctgcctaaactgtgc	633	
Qy	799	gtcataaccaaatcatttcatatttctaaacctcaaacaaacaaagctgtgtaatatctga	858	
Db	634	gtcataaccaaatcatttcatatttctaaacctcaaacaaacaaagctgtgtaatatctga	693	
Qy	859	tctctacggttccctcttggggcccaacattctccatatataccagccacactcatattt	918	
Db	694	tctctacggttccctcttggggcccaacattctccatatataccagccacactcatattt	753	
Qy	919	atttagttccagatctgtactgtgaacctttctacactgtagaataacattactatttt	978	
Db	754	atttagttccagatctgtactgtgaacctttctacactgtagaataacattactatttt	813	
Qy	979	gtcaagacccttcgtgtgtgctgcctaaatgtagctgactgttttccctaaaggagtg	1038	
Db	814	gtcaagacccttcgtgtgtgctgcctaaatgtagctgactgttttccctaaaggagtg	873	
Qy	1039	tctggccacggggatctgtgaacaggctgggaagcatctcaagatcttccagggttata	1098	
Db	874	tctggccacggggatctgtgaacaggctgggaagcatctcaagatcttccagggttata	933	
Qy	1099	cttactagcacacagcatgataattacaggagtgaaattatctaatcaaatatctcagtt	1158	
Db	934	cttactagcacacagcatgataattacaggagtgaaattatctaatcaaatatctcagtt	993	
Qy	1159	gtctttgccatactgaaatcatttccacttttgccacttttgccacttctcaagacctcaaat	1218	
Db	994	gtctttgccatactgaaatcatttccacttttgccacttttgccacttctcaagacctcaaat	1053	
Qy	1219	gtcattccattaaatacacaggattaaacttttttttaacctggagaagaattcaatgttta	1278	

Db	1054	gcaattccattaatatacaacagattaaacttttttttttaacctggagaagaattcaatgttta	1111
Qy	1279	catgcagctatgggaatttaattacataattttgttttccagtgcaaaagtactaaagtc	1338
Db	1114	catgcagctatgggaatttaattacataattttgttttccagtgcaaaagtactaaagtc	1173
Qy	1339	tttatccctccctttgtttgtatttttttccagatataaaagttaaaatgcttagccttgt	1398
Db	1174	tttatccctccctttgtttgtatttttttccagatataaaagttaaaatgcttagccttgt	1233
Qy	1399	actgagcgtgtatcacagcacagcctctcccacccctccacagccttatctgtcatcacact	1458
Db	1234	actgagcgtgtatcacagcacagcctctcccacccctccacagccttatctgtcatcacact	1293
Qy	1459	caacccctcccatnysacctaaacaaaactaaacttgtaattcccttgaaatgtcaggnc	1518
Db	1294	caacccctcccatnysacctaaacaaaactaaacttgtaattcccttgaaatgtcaggnc	1353
Qy	1519	atacatrttccctctgcctgagaagcctcttccctgtctcttaanctagaaatgatgtaa	1578
Db	1354	atacatrttccctctgcctgagaagcctcttccctgtctcttaanctagaaatgatgtaa	1413
Qy	1579	agttttgaataaagttgactatcttaactcatgcataaagaaggacacatatgagattcaatc	1638
Db	1414	agttttgaataaagttgactatcttaactcatgcataaagaaggacacatatgagattcaatc	1473
Qy	1639	atcacatlgagacagcaaaactactaaaagtgaatttgattataagagtttagataaaatata	1698
Db	1474	atcacatlgagacagcaaaactactaaaagtgaatttgattataagagtttagataaaatata	1533
Qy	1699	tgaatgcgaagakccacagaggaagtgttttatggggcacgcttttgtaagcctggagtgta	1758
Db	1534	tgaatgcgaagakccacagaggaagtgttttatggggcacgcttttgtaagcctggagtgta	1593
Qy	1759	agmaaaggcagggaacctcatagatctctatataataatacttcatcttctctatctctatc	1818
Db	1594	agmaaaggcagggaacctcatagatctctatataataatacttcatcttctctatctctatc	1653
Qy	1819	acaatatcccaacaagcttttccacagaattcatcgagtgcaaatccccaaaggtaaccttt	1878
Db	1654	acaatatcccaacaagcttttccacagaattcatcgagtgcaaatccccaaaggtaaccttt	1713
Qy	1879	atccatttcatgttgagtcgcgttttagaattttggcaaatcatctggtcacttatctca	1938
Db	1714	atccatttcatgttgagtcgcgttttagaattttggcaaatcatctggtcacttatctca	1773
Qy	1939	actttgagatgtgtttgtccctgtagttaattgaaagaaataggcactctttgtgagcca	1998
Db	1774	actttgagatgtgtttgtccctgtagttaattgaaagaaataggcactctttgtgagcca	1833
Qy	1999	ctttagaggttcaactcctcggcaataaagaatttacaaga	2037
Db	1834	ctttagaggttcaactcctcggcaataaagaatttacaaga	1872
RESULT 4			
AAA06687			
ID	AAA06687 standard; cDNA; 3112 BP.		
AC	AAA06687;		
XX			
XX			
DT	13-JUN-2000 (first entry)		
XX			
DE	Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:468.		
XX			
KW	Human; prostate cancer; diagnosis; tumour; gene therapy; detection;		
KW	immunogenic; cytostatic; vaccine; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	W0200004149-A2.		
XX			
PD	27-JAN-2000.		

QY 1811 tctctatcacatataccaaagctttttcacagaattcatgctgagtcagaaatccccaaagg 1870
|||||
Db 2860 tctctatcacatataccaaagctttttcacagaattcatgctgagtcagaaatccccaaagg 2919
|||||
QY 1871 taacctttatccatttcattggtgagtcgcttttagaatttttggcaaatcactactggtcac 1930
|||||
Db 2920 taacctttatccatttcattggtgagtcgcttttagaatttttggcaaatcactactggtcac 2979
|||||
QY 1931 ttatctcaactttgagatgtgttttctctttagttaaattgaaagaaataggcaactctt 1990
|||||
Db 2980 ttatctcaactttgagatgtgttttctctttagttaaattgaaagaaataggcaactctt 3039
|||||
QY 1991 gtgagccactttgaggttcaactccctggcaataaagaattttcaaaaga 2037
|||||
Db 3040 gtgagccactttgaggttcaactccctggcaataaagaattttcaaaaga 3086
|||||
RESULT 5
AAA06689/c
ID AAA06689 standard; cDNA; 2426 BP.
XX
AC AAA06689;
XX
DT
XX
XX 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:470.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
XX 27-JAN-2000.
PD
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 261-262; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumor protein (ptp). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAA82000 to AAA82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 2426 BP; 717 A; 476 C; 548 G; 685 T; 0 other;

Query Match 85.0%; Score 1731.4; DB 21; Length 2426;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1762; Conservative 4; Mismatches 2; Indels 3; Gaps 3;
QY 263 ggtgagaataagaagcgtgctgactttaccatcttgaggccacacatctgctgaaatgg 322
|||||
Db 1769 GGTGAGAAATAGAAAGGCTGCTGACTTTACCATCTGAGCCACACATCTGCTGAATGG 1710
|||||
QY 323 agataataatcaatcactagaacagacagatgacaataataatgtctaaagttagtgcacatgt 382
|||||
Db 1709 AGATAATTAACATCACTAGAAAACAGCAAGATGACAATAATAATGTCTAAGTAGTGCATGT 1650
|||||
QY 383 ttttgacattttccagcccttttaaatatccacacacagacagaaagaaaggaagcac 442
|||||
Db 1649 TTTTGACATTTCCAGCCCTTTAATAATATCCACACACAGGAAGCACAAGGAAGACAC 1590
|||||
QY 443 agagatccctgggagaaaatgcccggccgcacatcttgggtcactcgatgagcctcgccctgt 502
|||||
Db 1589 AGAGATCCTTGGGAGAAATGCCCGCGCCCATCTTTGGGTTCATCGATGAGCCTCGCCCTGT 1530
|||||
QY 503 gctgtgtcccgcttgtgaggggaagacattagaaaaatgaattgactgtctcccttaaaaga 562
|||||
Db 1529 GCTGTGTCCTCCGCTTGTGAGGGAAGGACATTAGAAAATGAATGTGTTCTCTTAAGGA 1470
|||||
QY 563 tgggcagaaaaacagatcctgttgggtatttttgaacgggattacagatttgaat 622
|||||
Db 1469 TGGGAGGAAAAACAGATCCTGTGTGGATATTTATTGAACGGGATTACAGATTTGAAAT 1410
|||||
QY 623 gaagtcaaaaagtgcattaccatgagaggaaaacacagacagaaaaacttcttgatggctt 682
|||||
Db 1409 GAAGTCACAAAGTGAGCATTTACCAATGAGGGAACACAGCAAGAAATCTTTGATGGCTT 1350
|||||
QY 683 cacaagacatgcaacaacaaaatggaaatactgtgatgacatgagggagcagccaagtgggg 742
|||||
Db 1349 CACAAGACATGCAACAAACAAAATGGAATPACTGTGTATGATGACATGAGCAGCAAGCTGGGG 1290
|||||
QY 743 aggaataaacaaggggagaggggtcaggattctggccctgctgctaaactgtcggttc 802
|||||
Db 1289 AGGATAAACCCAGGGGACAGGGGTACAGGATTTCTGGCCCTGCTGCTAAACTGTGGGTTT 1230
|||||
QY 803 ataaccacaaatcatcttcaatttttaacctcaaaacaaagctgtgtgtaataatctgacttc 862
|||||
Db 1229 ATAACCAAAATCATTTTCATATTTCTAACCCCTCAAAACAAAGCTGTGTATATCTGATCTC 1170
|||||
QY 863 tacggttcctctgtgggcccacacattctccatatatccagccacactcaatttttaatttt 922
|||||
Db 1169 TACGGTTCTTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCATTTTAAATATTT 1110
|||||
QY 923 agttccagatctgactgtgacctttctacactgtagaataacattacttttttttc 982
|||||
Db 1109 AGTTCCAGATCTGACTGTGACCTTTCTACACTGTAGAAATACATTAATCTCAATTTTGTTC 1050
|||||
QY 983 aaagacccttctgtgtgctgctaaatagtactgactgtttttccaaaggaggttcttg 1042
|||||
Db 1049 AAAGACCTTCTGTTGCTGCTAAATATATGTAGTGAATGCTTTTCTTAAGGAGTGTCTG 990
|||||
QY 1043 gccaggggatctgtgaacaggtgggaagcatctccaaagattttccagggttatactta 1102
|||||
Db 989 GCCCAGGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGGTTATACTTFA 930
|||||
QY 1103 ctgacacacagatcatcattacggagtgaaattatataataacatcattcctcagttgtct 1162
|||||
Db 929 CTAGCACACAGGATGATCATTTACGGAGTGAATTAATTAATCAACATCATCTCCAGTGTCT 870
|||||
QY 1163 ttgcccactactgaaatttcatttccacctttgtgcccatttccagacctcaaaatgtca 1222
|||||
Db 869 TTGCCCATACTGAATTCATTTCCCACTTTTGTGCCCACTTCTCAAGACCTCAAAATGTCA 810
|||||
QY 1223 ttccattaaatcacaggattaaacttttttttttaacctgggaagaattcaattgttacatg 1282
|||||
Db 809 TTCCATTAAATATCACAGGATTAACTTTTTTTTTTTTAACTTGAAGAATTAATGTTTACATG 750
|||||

QY	1283	cagctatgggaatttaattacaataatttggtttccagtgcaaaagatgactaaagtccttta	1342
Db	749	CAGCTATGGGAATTTAAATTACATAATTTGTTTCCAGTGCAAAAGATGACTAAAGTCCTTTA	690
QY	1343	tcctccctctgttggattttttccagftaanaagttaaatgcttagccttgactg	1402
Db	689	TCCTCCCTTGTTGGATTTTTTTCAGTATAAAGTTAAATGCTTAAGCTTGACTG	630
QY	1403	aggctgtatcacg - cacagcctctcccatccctccagccttatctgtcatcaccatcaa	1461
Db	629	AGGCTGTATACAGCCACAGCCTCTCCCATCCCTCCAGCCTTATCTGTCAATCACCATCAA	570
QY	1462	ccctcccatnysacctaaacaaatactaaactgtaattccttgaaacatgtcaggncata	1521
Db	569	CCCTCCCATG - CACCTAAACAAAATCTAACTGTGTAATCTCTGAAATGTCTAGG - CATA	512
QY	1522	cattttcctctgcctgagaagctcttcctgtctcttaantctagaatgagtgaagt	1581
Db	511	CATTATTCTCTGCCTGAGAAGCTTCTCCTGTCTCTTAATCTAGAAATGATGAAAGT	452
QY	1582	tttgaataagttgactatcttacttcatgcгааааgggacacatatgagatttcatactc	1641
Db	451	TTTGAATAAGTTGACTTATCTTCTTCATGCAAAGAGGACACATATGAGATTCAATCATC	392
QY	1642	acatgagacagcaaatactaaaagtgtaatttgattgattataagatttagataaataatga	1701
Db	391	ACATGAGACAGCAAAATCTPAAAAGTGTAATTTGATTATAAGAGTTTAGATAAATATATGA	332
QY	1702	aatgcaagakccacagagggaaatgtttatggggcacgtttgtgaagcctggagtgtaagm	1761
Db	331	AAATGCAAGAGCCACAGAGGGAATGTTATGGGGACGTTTGTAAGCCTGGGATGTGAAGC	272
QY	1762	aaaggcaggaaacctcatgattctataataataacttcattctctatctctatacaaa	1821
Db	271	AAAGGCAGGGAACCTCATAGTATCTTATATAATATACPTTCATTTCTCTATCTATACACA	212
QY	1822	atatccaaacaagcttttcagaaatctcatgagtgcaaatccccaaaggtaaacctttatc	1881
Db	211	ATATCCAACAAGCTTTTCACAGAATTCATGCAGTGCAAATCCCCAAAGGTAACCTTTATC	152
QY	1882	catttcattggtgagtgcgcttagaattttggcaaatcatctactggtcacattatctcaact	1941
Db	151	CATTTCATGGTCAGTCGCGCTTTAGAAATTTTGGCAAAATCATACTCGTCACATTATCTCAACT	92
QY	1942	ttgagatggtttgttccttgtagttaaattgaaagaaataggggcactcttgtgagccaatt	2001
Db	91	TTGAGATGTTTGCTCTGTGTAGTTAAATGAAAGNAATAGGGCACTCTTGTGAGCCACTT	32
QY	2002	tagggttcactcctggcaataaagaatttac	2032
Db	31	TAGGGTTTCACTCTCTGGCAATAAAGAAATTTAC	1

RESULT	6	
AAA06688/C		
ID	AAA06688	standard: cdNA; 2229 BP.
XX		
XX	AAA06688;	
XX		
DT	13-JUN-2000	(first entry)
XX		
DE	Human immunogenic prostate tumour protein cdNA sequence	SEQ ID NO:469.
XX		
KW	Human; prostate cancer; diagnosis; tumour; gene therapy; detection;	
KW	immunogenic; cytostatic; vaccine; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200004149-A2.	
XX		
PD	27-JAN-2000.	
XX		

Pf	14-JUL-1999;	99WO-USL5838.
XX		
PR	14-JUL-1998;	98US-0115453.
PR	14-JUL-1998;	98US-0116134.
PR	23-SEP-1998;	98US-0159812.
PR	23-SEP-1998;	98US-0159822.
PR	15-JAN-1999;	99US-0232149.
PR	15-JAN-1999;	99US-0232880.
PR	09-APR-1999;	99US-0288946.
XX		
XX	(CORI-) CORIXA CORP.	
PA		
XX	Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;	
PI		
XX	WPI; 2000-171268/15.	
DR		
XX		
PT	New polypeptide useful for treating and diagnosing prostate cancer	
PT	comprises an immunogenic portion of prostate tumor protein -	
XX		
PS	Claim 1; Page 260-261; 263pp; English.	
XX		
CC	The present invention describes isolated polypeptides, comprising an	
CC	immunogenic portion of a prostate tumour protein (PrP). The polypeptides	
CC	and polynucleotides encoding them have cytostatic activity and can be	
CC	used in vaccines and in gene therapy. The polypeptides and	
CC	polynucleotides encoding them, antigen presenting cells which express	
CC	the polypeptides, antibodies against the polypeptides and vaccines	
CC	comprising them can be used for inhibiting the development of prostate	
CC	cancer in a patient. The polypeptides can be used to generate antibodies	
CC	or anti-idiotypic antibodies for passive immuno therapy. A portion of	
CC	the polynucleotides encoding the polypeptides can be used as a probe or	
CC	to modulate the expression of the polypeptides. AAA06241 to AAA06691 and	
CC	AAAY82000 to AAAY82020 represent sequences used in the exemplification of	
CC	the present invention.	
XX		
SQ	Sequence 2229 BP; 654 A; 447 C; 481 G; 647 T; 0 other;	
Query Match 84.1%; Score 1712.4; DB 21; Length 2229;		
Best Local Similarity 99.4%; Pred. No. 0;		
Matches 1765; Conservative 4; Mismatches 2; Indels 5; Gaps 5;		
QY	263 ggtgagaataagaaggctgcgtgaactttaccatcttgagccacacacatcgtcgaaatgg	322
Db	1775 GGTCAGAAATTAAGAAAGCGCTGCTGACTTTACATCTGAGGCCACACATCTGTCAAATGG	1716
QY	323 agataattcaatcacatagaaaacagcaagatgacaatatataatgtctaagttagtcacatgt	382
Db	1715 AGATAATTAACTCATCTAGAAAACAGCAAGATGACATATATATGCTTAAAGTAGTGACATGT	1656
QY	383 ttttgcacatttcagccccctttaaatatccacacacacaggaagcaaaaaagaaagcac	442
Db	1655 TTTTGGCACATTTCCAGCCCCCTTTAAATATCCACACACAGGAAGCACAAGAAGGCAC	1596
QY	443 agagatcccctggagaaaatgcccgcccatcttgggtcatcgatgagcctgcacctgt	502
Db	1595 AGAGATCCCCTGGAGAAATGCCCGCCGCATCTTGGGTTCATCGATGAGGCTCGCCCTGT	1536
QY	503 gcctggctcccgcctgtgagggaaaggacattagaaaaatgaattgatgtgttcctcaaagga	562
Db	1535 GCCTGGTCCCGCTTGTGAGGGAGGACATTTAGAAAAATGAATTGATGTGTTCTTAAAGGA	1476
QY	563 tgggcaggaaaaacagatccctgttgtggatatttatattgaacgggattcacagatttgaat	622
Db	1475 TGSGCAGAAAACAGATCCCTGTGTGTGGATATTTATTTGAACGGGATTTACAGATTTGAAAT	1416
QY	623 gaggtcaaaagtggagcattaccaatgagggaaaaacagacgagaaaaatcttgatgctt	682
Db	1415 GAAGTCACAAGGTGAGCTTTACCATGAGAGGAAAACAGACGAGAAAATCTTGTATGCTT	1356
QY	683 cacaagacatgcacaacaaaaatggaatactgtgatgatgagcgagcgaactgggg	742
Db	1355 CACAACAGCATGCAACAAACAAATGGAATCTGTGTGATGACATGAGCAGCAAGCTGGGG	1296

Db	364	AAATGAAGTCACAAAGTGAGCATTTACCAATGAGAGGAAACAGACAGGAGAAAATCTTGATG	305
Qy	679	gcttcacaagacatgcaacaaacaaatggaatactgtgatgacatgaggcagccaagct	738
Db	304	GCCTTCAAGACATGCAACAAACAAATGGAATACTGTGATGACATGAGGCAGCCAAGCT	245
Qy	739	ggggaggaataaacacaggggcagaggggtcaggaattctggccctgctgcctaaactgtgc	798
Db	244	GGGGAGGAGATAACACGGGGCAGAGGGTCAAGGATTCTGGCCCTGCTGCCCTAAACTGTGC	185
Qy	799	gtctacaacaaatcatcttcataatttctaaccctcaaaacaaagctgttctaatactga	858
Db	184	GTTTCATACCAAAATCATTTTCATATTTCTTACCTCAAAACAAAGCTGTTGTAATCTCGA	125
Qy	859	tctctacggttctctctggggcccaacattctccatatatccacgcacacactcatTTtaat	918
Db	124	TCCTACGGTCTCTTCTGGGCCCAACATTTCTCCATATATCCAGGCACACTCATTTTTAAT	65
Qy	919	atttaattccagatctgtactgtgaacatttctacactgtagaataaacattactcatTTt	978
Db	64	ATTTAGTCTCCAGATCTGTACTGTGACCTTTCTACACTGTAGAATAACATTACTCATTTT	5
Qy	979	gttc 982	
Db	4	GTTC 1	
RESULT	9		
AAX37486			
ID	AAX37486	standard; cDNA; 597 BP.	
XX			
AC	AAX37486;		
XX			
DT	06-JUL-1999	(first entry)	
XX			
DE	Human secreted protein cDNA fragment containing gene 36.		
XX			
KW	Human; secreted protein; treatment; prevention; protein therapy; AIDS;		
KW	gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;		
KW	developmental abnormality; fetal deficiency; blood disorder; leukemia;		
KW	immune system disease; autoimmune disease; hepatic disease; lymphoma;		
KW	renal disease; inflammation; allergy; Alzheimer's disease; schizoprenia;		
KW	cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;		
KW	pulmonary disorder; transplant rejection; osteoclast; osteoporosis;		
KW	arthritis; malignancy; digestive; endocrine; infection; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	W09918208-A1.		
XX			
PD	15-APR-1999.		
XX			
PF	01-OCT-1998; 98WO-US20775.		
XX			
PR	02-OCT-1997; 97US-0060884.		
PR	02-OCT-1997; 97US-0060833.		
PR	02-OCT-1997; 97US-0060836.		
PR	02-OCT-1997; 97US-0060837.		
PR	02-OCT-1997; 97US-0060838.		
PR	02-OCT-1997; 97US-0060839.		
PR	02-OCT-1997; 97US-0060843.		
PR	02-OCT-1997; 97US-0060862.		
PR	02-OCT-1997; 97US-0060866.		
PR	02-OCT-1997; 97US-0060874.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;		
PI	Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;		
PI	Rosen CA, Ruben SM, Shi Y, Young P, Yu G;		
XX			
DR	WPI; 1999-264022/22.		
DR	P-PSDB; AAY07887.		

Query Match	15.7%;	Score 319.6;	DB 21;	Length 437;
Best Local Similarity	98.5%;	Pred. No. 5.3e-82;		
Matches 319;	Conservative	2;	Mismatches	3; Indels 0; Gaps 0;

QY	259	caggggtagaataaagaaagcgtgctgactttaccatctgagccacacatctgctgaa	318
DB	114	cagaggtagaataaagaaagcgtgctgactttaccatctgagccacacatctgctgaa	173
QY	319	atggagataataacatcactagaaaacagcgaagatgacaataataatgtctaagtgtgac	378
DB	174	atggagataataacatcactagaaaacagcgaagatgacaataataatgtctaagtgtgac	233
QY	379	atgtttttgacatttcagccctttaataatccacacacacaggaagcacaagaa	438
DB	234	atgtttttgacatttcagccctttaataatccacacacacaggaagcacaagaa	293
QY	439	gcacagagatccctggagaaaatgcccgccgcctcttgggtcatgcatgagcctgccc	498
DB	294	gcacagagatccctggagaaaatgcccgccgcctcttgggtcatgcatgagcctgccc	353
QY	499	ctgtgcctggtcccgctgtgaggaaggaacattagaaaaatgaattgatgtgttccttaa	558
DB	354	ctgtgcctggtcccgctgtgaggaaggaacattagaaaaatgrattgatgtgttccttaa	413
QY	559	aggatgggcaggaaaacagatcct	582
DB	414	aggatgggcaggaaaacagatcct	437

RESULT 12	
AAA06520/c	
ID AAA06520 standard; cDNA; 301 BP.	
XX	
AC AAA06520;	
XX	
DT 13-JUN-2000 (first entry)	
XX	
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:287.	
XX	
KW *Human; prostate cancer; diagnosis; tumour; gene therapy; detection;	
KW immunogenic; cytostatic; vaccine; ss.	
XX	
OS Homo sapiens.	
XX	
PN W0200004149-A2.	
XX	
PD 27-JAN-2000.	
XX	
XX 14-JUL-1999; 99WO-US15838.	
XX	
PR 14-JUL-1998; 98US-0115453.	
PR 14-JUL-1998; 98US-0116134.	
PR 23-SEP-1998; 98US-0159812.	
PR 23-SEP-1998; 98US-0159822.	
PR 15-JAN-1999; 98US-0232149.	
PR 15-JAN-1999; 99US-0232880.	
PR 09-APR-1999; 99US-0288946.	
XX	
XX (CORI-) CORIXA CORP.	
XX	
PI Dillion DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;	
XX	
XX WPI; 2000-171268/15.	
DR	
XX	
PT New polypeptide useful for treating and diagnosing prostate cancer	
PT comprises an immunogenic portion of prostate tumor protein -	
XX	
XX	
PS Claim 1; Page 192; 263pp; English.	
XX	
CC The present invention describes isolated polypeptides, comprising an	
CC immunogenic portion of a prostate tumour protein (ptp). The polypeptides	
CC and polynucleotides encoding them have cytostatic activity and can be	

[illegible]


```

RESULT 15
AAF58254/C
ID AAF58254 standard; DNA; 936 BP.
XX
XX AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA Umek RM;
PI WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
PT
XX Example 6; Page 127; 159pp; English.
PS
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match          6.7%; Score 136; DB 22; Length 936;
Best Local Similarity 0.9%; Pred.NO. 5.4e-29;
Matches 7; Conservative 491; Mismatches 279; Indels 0; Gaps

Qy 1132 aattctatcaacatcatcctcagtgctttggccataactgaaatttcacctt 1191
      : :: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 777 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 718

Qy 1192 ttgtgccattctcaagacctcaaatgtcattcattaatacacaggattaacctttt 1251
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 717 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 658

Qy 1252 tttttaacctggaagaattcaatgtttacatgcagctatgggaatttaatatcatatttg 1311
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 657 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 598

Qy 1312 ttttcagtgcacaagatgactaagtcctttatccctccccttgttgattttttcca 1371
      : : : | : : : : : : : : : : : : : : : : : : : : : : :
Db 597 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 538

Qy 1372 gtataaggttaaaatgcttagcctgtactgagcgctgtatacacgacagcctctccccat 1431
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 537 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 478

Qy 1432 cccctccagccttatctgtcatcacatcaacccctcccatnysaccttaacaaaatactaa 1491

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 01:48:13 ; Search time 10902.7 Seconds
(without alignments)
2889.911 Million cell updates/sec

Title: US-09-402-713A-1
Perfect score: 2037
Sequence: 1 aagaagctggcatcagaaaa.....caataaagaattacaaga 2037

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_bal:
2: gb_ba2:
3: gb_ba3:
4: gb_in1:
5: gb_in2:
6: gb_in3:
7: gb_om:
8: gb_ov:
9: gb_pat1:
10: gb_pat2:
11: gb_ph:
12: gb_pl1:
13: gb_pl2:
14: gb_pl3:
15: gb_pl4:
16: em_ba1:
17: em_ba2:
18: em_fun:
19: em_htgo_hum:
20: em_htgo_inv:
21: em_htgo_rod:
22: em_htg_hum1:
23: em_htg_hum2:
24: em_htg_hum3:
25: em_htg_hum4:
26: em_htg_hum5:
27: em_htg_hum6:
28: em_htg_hum7:
29: em_htg_hum8:
30: em_htg_inv1:
31: em_htg_inv2:
32: em_htg_inv3:
33: em_htg_rod:
34: em_hum1:
35: em_hum2:
36: em_hum3:
37: em_hum4:
38: em_hum5:
39: em_hum6:
40: em_hum7:
41: em_in:
42: em_om:
43: em_or:

44: em_ov:
45: em_pat:
46: em_ph:
47: em_pl:
48: em_ro:
49: em_sts:
50: em_sy:
51: em_un:
52: em_vi:
53: gb_sts1:
54: gb_sts2:
55: gb_sts3:
56: gb_sy:
57: gb_un:
58: gb_vil:
59: gb_vl2:
60: gb_vl3:
61: gb_vl4:
62: gb_vl5:
63: gb_vl6:
64: gb_vl7:
65: gb_vl8:
66: gb_vl9:
67: gb_vl10:
68: gb_vl11:
69: gb_vl12:
70: gb_vl13:
71: gb_vl14:
72: gb_vl15:
73: gb_vl16:
74: gb_vl17:
75: gb_vl18:
76: gb_vl19:
77: gb_vl20:
78: gb_vl21:
79: gb_vl22:
80: gb_vl23:
81: gb_vl24:
82: gb_vl25:
83: gb_vl26:
84: gb_vl27:
85: gb_vl28:
86: gb_vl29:
87: gb_vl30:
88: gb_vl31:
89: gb_vl32:
90: gb_vl33:
91: gb_vl34:
92: gb_vl35:
93: gb_vl36:
94: gb_vl37:
95: gb_vl38:
96: gb_vl39:
97: gb_vl40:
98: em_ba3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2032	99.8	3923	88	AF103907	Homo sapi
2	1589.4	78.0	5435	88	AF103908	Homo sapi
3	1555.8	76.4	164371	80	AL390239	Homo sapi
4	1361.4	66.8	173831	80	AL359314	Homo sapi
5	1317.8	64.7	267581	80	AL358573	Homo sapi
6	235.4	11.6	359	9	AX018075	Sequence
7	99	4.9	143675	79	AL161625	Homo sapi
8	99	4.9	267581	80	AL358573	Homo sapi

```
9      98      4.8 580 89 AF279290 Homo sapi
c 10 97.4 4.8 172298 61 AC009556
11 59.8 2.9 7218 10 I66494
c 12 57.4 2.8 7218 10 I66494
13 52.6 2.6 1141 10 AX083744
14 47.2 2.3 137064 74 AC073363
c 15 47.2 2.3 152571 66 AC021589
c 16 47.2 2.3 160133 69 AC025140
c 17 47.2 2.3 167574 62 AC011274
c 18 47.2 2.3 177396 74 AC069219
c 19 47.2 2.3 189315 70 AC028861
c 20 47.2 2.3 191034 80 AL356372
c 21 47 2.3 10968 9 AR029638
c 22 47 2.3 10968 14 LSU65391
c 23 47 2.3 59202 13 AF220603
c 24 47 2.3 60831 13 AF220602
c 25 45.8 2.2 188781 67 AC022546
c 26 45.8 2.2 323479 77 AC087779
c 27 45.6 2.2 40873 83 CEH04109
c 28 45.6 2.2 42141 6 CEY54G9A
c 29 45.6 2.2 300197 83 CEY54G9
c 30 45.4 2.2 5134 9 AR085492
c 31 45.4 2.2 5475 9 AR029637
c 32 45.2 2.2 1141 10 AX083744
c 33 45 2.2 191319 63 AC015653
c 34 44.4 2.2 151553 90 AL139342
c 35 44.4 2.2 181545 69 AC025529
c 36 44.2 2.2 163958 4 AC010705
c 37 44 2.2 106935 6 CEY137E3
c 38 43.8 2.2 171471 64 AC015844
c 39 43.2 2.1 120984 86 AC006463
c 40 43.2 2.1 184316 68 AC024022
c 41 42.8 2.1 3269 2 AF270032
c 42 42.8 2.1 9131 85 AB038781
c 43 42.6 2.1 930 53 CNS08TZZ
c 44 42.6 2.1 8912 85 AB038782
c 45 42.6 2.1 10708 89 AF113616
```

ALIGNMENTS

```
RESULT 1
AF103907 AF103907 3923 bp mRNA PRI 14-AUG-2000
LOCUS Homo sapiens non-coding RNA DD3 sequence.
ACCESSION AF103907
VERSION AF103907.1 GI:6165973
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3923)
AUTHORS Bussemakers,M.J., van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaus,H.F., Schalken,J.A., Debruyne,F.M., Ru,N. and Isaacs,W.B.
DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072260
PUBMED 10606244
REFERENCE 2 (bases 1 to 3923)
AUTHORS Bussemakers,M.J.G., van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaus,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru,N. and
Isaacs,W.B.
Direct Submission
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
LOCATION/Qualifiers
FEATURES
1..3923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
```

```
/map="9q21-q22"
/tissue_type="prostate"
/note="non-coding RNA DD3"
repeat_region 3756..3911
/note="LINE"
/rpt_family="L1"
/rpt_type=dispersed
BASE COUNT 1157 a 840 c 740 g 1186 t
ORIGIN
Query Match 99.8%; Score 2032; DB 88; Length 3923;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2029; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
QY 1 aagaactgcatacaaaaaaacagagggagattgtgtgtgtgcagccagggagaccag 60
Db 23 AGAAGCTGGCATCAGAAAAACAGAGGGGAGATTGTGTGGCTGCAGCCGAGGGAGCCAG 82
QY 61 gaagatctgcatggggaaagaccctgattgatacagagaattacaacacatatcttag 120
Db 83 GAAGATCTCATGGTGGGAAGGACCTGATGATACAGAGGAATTACAACACATATACTTAG 142
QY 121 ttttcaatgaacacacacacataaataagtgaagagctagtcgcgtgtgagtcctcctcag 180
Db 143 TGTTCATGAACACCAACAGATAAATAAGTGAAGAGCTAGTCCGCTGTGAGTCTCCTCAGT 202
QY 181 gacacagggctgataccatcgacgggacatttctgagtactcaagtgcagcaaaaagaa 240
Db 203 GACACAGGCTGGATCACCATCGACGGCACTTCTGAGTACTCAGTGCAGCAAAAGAAAG 262
QY 241 ctacagacatctcaatggcaggggagaaataaagaaagcctgctgactttaccatctga 300
Db 263 CTACAGACATCTCAATGGCAGGGGTGAGAAATAAGAAAGGCTGCTGACTTTTACCATCTGA 322
QY 301 gggccacacatctgtgaaatggagataattaacatcactagaacacagcaagatgacaata 360
Db 323 GGCCACACATCTGCTGAATGGAGATAATTAACATCCTAGAAACAGCAAGATGACAATA 382
QY 361 taatgtctaagttagtgcacatgtttttgcacattccagccctttaataatccacacaca 420
Db 383 TAATGCTCTAAGTAGTGACATGTTTTTGGACATTTCCAGCCCTTTTAAATATATCCACACA 442
QY 421 caggaagcacaaaaaagaacacagagatccctgggagaaatcccgggccgcctcttggg 480
Db 443 CAGGAGACACAAAAGGAAGACACAGAGATCCCTGGGAGAAATGCCCGCCGCCATCTTGGG 502
QY 481 tcatgatgagcctgcctgtgcctgggtcccgcttgtgaggaaggaacattagaaaaatg 540
Db 503 TCATCGATGAGCCTCGCCCTGTGCCTGGTCCCGCTTGTGAGGGAAGGACATTAGAAAAATG 562
QY 541 aattgatgttcttaaggatgggagaaaaacagatccctgtgtgtgataattattttg 600
Db 563 AATTGATGTGTTCTTAAAGGATGGGCAGGAAACAGAGATCCTGTGTGGATATTATTATTG 622
QY 601 aacggattacagattgaaatgaagtcaacaagtgagcattaccaatgagggaaaaaca 660
Db 623 AACGGGATTACAGATTGAAATGAAGTCAACAAGTGAGCATTTACCATGAGAGGAAAAACA 682
QY 661 gacgagaaaaattgtatggcttcacaagacatgcacaaacaaaatggaaatgtgatg 720
Db 683 GACGAGAAATCTTGATGGCTTCAACAAGACATGCAACAAACAAATGGAATACTGTGATG 742
QY 721 acatgaggcagccaagctggggaggagataaacacggggcagagggtaaggattctggcc 780
Db 743 ACATGAGGAGCAACAGCTGGGGAGGAGATAACACGGGGCAGAGGCTCAGGATTTCTGGCC 802
QY 781 ctgctgcttaactgtgcgttcataacaaaatcatttcatttcttaaccctcaaaaaca 840
Db 803 CTGCTGCCCTAAACTGTGCGTTCATACCAAAATCATTTTCATATTCTTAACCCCTCAAAACA 862
QY 841 agctgtgtaatatctgatctctacggttctcttcttcttcttcttcttcttcttcttcc 900
```

Db 863 AGCTGTTGTAATATCTGATCTCTACGGTTCCTTCGGGCCCAACATTCCTCCATATATCCA 922
Qy 901 gccacactcattttaatttagttccagatctgtactgtgaccttctacactgtag 960
Db 923 GCCACACTCATTTTAAATATTTAGTTCCAGATCTGTACTGTGACCTTTCTACACTGTAG 982
Qy 961 aatacattactcatttggttcaagacccttcgtgtgctgcctaaatagtactgact 1020
Db 983 AATAACATTACTCATTTGTTCAAGACCCCTCGTGTGCGCTAAATATGTAGCTGACT 1042
Qy 1021 gtttttccaaaggagtggtctggccagggatctgtgaacaggctgggaagcatctcaa 1080
Db 1043 GTTTTTCCTCAAGGAGTGTCTGGCCAGGGATCTGTCAACAGGCTGGGAAGCATCTCAA 1102
Qy 1081 gatcttccagggttatacttactacagacacagcatgatacttaccgagtgaaattatcta 1140
Db 1103 GATCTTTCCAGGGTTATCTTACTAGCACAGCATGATCATTCGGGAGTGAATTTATCTA 1162
Qy 1141 atcaacatcctcagtgctcttggccatactgaaattcatttccacttttggccca 1200
Db 1163 ATCACATCATCTCAGTGTCTTTGCCCATACTGAAATTCATTTCCCACTTTTGTGCCCA 1222
Qy 1201 ttctcaagacctcaaaatgtcattccatttaataatcacaggatgaacttttttttaacc 1260
Db 1223 TTCTCAAGACCTCAAAATGTCATTCATTAATATCACAGGATTAATTTTAAACC 1282
Qy 1261 tggagaataatcaatgtttacatgcagctatgggaatttaattacatatattgtttccagt 1320
Db 1283 TGGAGAATAATCAATGTTACATGCAGCTATGGGAATTTAATACATATTTGTTTCCAGT 1342
Qy 1321 gcaaatgactaaagcttctatccctccctttgttgatttttttccagtataaagt 1380
Db 1343 GCAAGATGACTAAGTCTTTATCCCTCCCTTGTGTTGATTTTTTCCAGTATAAAGT 1402
Qy 1381 taaatgtctgactgtactgagtgctgtatacagcacagcctctcccatccctccagc 1440
Db 1403 TAAATGCTTAGCCTTGTACTGAGGCTGTATACAGCACAGGCTCTCCCACTCCCTCCAGC 1462
Qy 1441 ctatctgtcatcaccatcaacccctcccatnysacctaaacaaatactactgttaatt 1500
Db 1463 CTTATCTGTCTATCACCATCAACCCCTCCCATACACCTTAACAAATCTAATCTGTAAT 1522
Qy 1501 ccttgaacatgtcaggncatacatrttccctctgctgagagctcttctctgtctctt 1560
Db 1523 CCTTGAACATGTCAGGACATACATTTTCCCTCTGCTGAGAGCTCTTCCCTGTCTCT 1582
Qy 1561 aantctagaatgatgaaagttttgaaataagttgactatcttacttcatgcaaaagagg 1620
Db 1583 AAATCTAGAATGATGTAAAGTTTGAATAAGTTGACTATCTTACTTCATGCAAGAAGGG 1642
Qy 1621 acacatatgagattcatcatcacatgacagcaaaataactaaaagttaattgtattata 1680
Db 1643 ACACATATGAGATTTCATCATCATGAGACAGCAAAATCTAAAGTGAATTTGATTATA 1702
Qy 1681 agagtttagataaataatatgaatacagkakccagaggaatgtttatggggcacgtt 1740
Db 1703 AGAGTTTAGATAAATATATGAATGCAGAGCCACAGAGGATGTTTATGGGGCACGTT 1762
Qy 1741 tgaagcctggagtgtagmaaaaggcagggaacctcatagtatcttatataataactt 1800
Db 1763 TGAAGCCTGGGATGTGAAGCAAGGACGGAACCTCATAGTATCTTATATAATATACTT 1822
Qy 1801 catttctctatctctacacaaatcccaagaagcttttcaagaattcatgcagtgcaaa 1860
Db 1823 CATTTCTCTATCTATCATATATCCAAACAGCTTTTTCACAGAATTCATGCAGTGCAAA 1882
Qy 1861 tcccaaaaggtaacctttatccatttcattggtgagtcgcgttagaattttggcaaatca 1920
Db 1883 TCCCAAAAGTAAACCTTTATCCATTTTCATGGTGGCTTTAGAAATTTGGCAATCA 1942
Qy 1921 tactggtcacttatctcaactttagatgtgttcttcccttagtttaattgaaagaaata 1980

Db 1943 TACTGGTCACCTTATCTCAACTTTGAGATGTGTTGCTCCTGTAGTTAATTTGAAAGAAATA 2002
Qy 1981 gggcactcttgagcaccatttaggtttcactcctgccaataaagaatttacaaga 2037
Db 2003 GGGCACTCTTGTGAGCCACTTTAGGGTTCACTCTCTGGCAATAAAGAATTTACAAGA 2059
RESULT 2
AF103908 Locus AF103908 5435 bp DNA PRI 14-AUG-2000
DEFINITION Homo sapiens non-coding RNA DD3 gene, exons 2, 3, and 4.
ACCESSION AF103908
VERSION AF103908.1 GI:6165974
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5435)
Bussemakers,M.J., van Bokhoven,A., Verhaegh,G.W., Smit,F.P., Karthaus,H.F., Schalken,J.A., Debruyne,F.M., Ru,N. and Isaacs,W.B.
DD3: a new prostate-specific gene, highly overexpressed in prostate cancer
TITLE
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072260
PUBMED 10606244
REFERENCE 2 (bases 1 to 5435)
AUTHORS Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P., Karthaus,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru,N. and Isaacs,W.B.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
source 1..5435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q21-q22"
repeat_region 1..78
/rpt_family="AluY"
/rpt_type="dispersed"
exon 533..697
/note="Alternative exon present in 5% of cDNA clones"
repeat_region 1035..1294
/rpt_family="Alu"
/rpt_type="dispersed"
mRNA join(1571..1753,1981..5435)
/note="transcript III"
mRNA /product="non-coding RNA DD3"
join(1571..1753,1981..2517)
/note="transcript I"
mRNA /product="non-coding RNA DD3"
join(1571..1753,1981..3579)
/note="transcript (major) II"
exon /product="non-coding RNA DD3"
1571..1753
/number=3
exon 1981..5435
/number=4
polyA_signal 2495..2499
/note="transcript I"
polyA_site 2517
/note="transcript I"
polyA_signal 3553..3558
/note="transcript (major) II"
polyA_site 3579
/note="transcript (major) II"
repeat_region 5268..5423
/note="LINE"
/rpt_family="L1"
/rpt_type="dispersed"

[illegible]

Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 163337 bases at least Q40
Consensus quality: 163481 bases at least Q30
Consensus quality: 163577 bases at least Q20
Insert size: 164071; sum-of-contigs
Insert size: 166918; 4.2% error; agarose-fp
Quality coverage: 8.78x in Q20 bases; sum-of-contigs Quality
coverage: 9.07x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 20240: contig of 20240 bp in length
* 20241 20340: gap of 100 bp
* 20341 116100: contig of 95760 bp in length
* 116101 116200: gap of 100 bp
* 116201 118243: contig of 2043 bp in length
* 118244 118343: gap of 100 bp
* 118344 164371: contig of 46028 bp in length.

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-58J3"
/clone.lib="RPC1-11.1"

misc_feature

1..20240
/note="assembly_fragment:00838
fragment_chain:1
clone_end:SP6
vector_side:left"

misc_feature

20341..116100
/note="assembly_fragment:01998
fragment_chain:1"

misc_feature

116201..118243
/note="assembly_fragment:00916
fragment_chain:1"

misc_feature

118344..164371
/note="assembly_fragment:00334
fragment_chain:1
clone_end:T7
vector_side:right"

BASE COUNT 50254 a 34905 c 32663 g 46249 t 300 others
ORIGIN

Query Match 76.4%; Score 1555.8; DB 80; Length 164371;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1587; Conservative 4; Mismatches 3; Indels 3; Gaps 3;
QY 442 cagagatccctggagaaatgccgcgcacatcttgggtcatcgatgagcctgccttg 501
DB 33316 CACAGATCCCTGGGAGAAATGCCGCGCCCATCTTGGGTCTATCGATGAGCCTGCCCTG 33375
QY 502 tgcctggtccctgtgaggaagacattagaaataaattgatgtgtctcttaaagg 561
DB 33376 TGCCTGGTCCCTGTTGAGGAGGACATAGAAAATGAATGATGTGCTCTTAAGG 33435
QY 562 atggcaggaacacagatcctgttggatattttgaacgggattacagattgaa 621
DB 33436 ATGGCAGGAAAACAGATCCTGTTGGGATATTTATTTGAACGGGATTACAGATTGAAA 33495
QY 622 tgaagtcaaaagttagcattccatagagagaaacacagacgaaatactttagcct 681
DB 33496 TGAAGTCAAAAAGTGAGCATTAACCAATGAGAGGAAACAGACGAGAAAATCTTGATGGCT 33555

QY 682 tcacaagacatgcaacaaacaaatggaatactgtgatgacatgaggcagccaaactggg 741
DB 33556 TCACAAAGACATGTCACAAACAAATGGAATACTGTGATGACATGAGGCAGCAAGCTGGG 33615
QY 742 gaggagataaacacacgggagagggtcaggattctggcctgtgctcctaaactgtgcgtt 801
DB 33616 GAGGAGATAACACACGGGCAGAGGCTCAGGATCTTGCCCTGCTGCTAAACTGTGCGGT 33675
QY 802 cataacaaaatcattctcatttctaaaccccaaaacaaagctgttgaatactgatct 861
DB 33676 CATACACAAATCATTTTCATATTTCTAACCCCTCAAAACAAAGCTGTGTAATATCTGATCT 33735
QY 862 ctacggttctctctgggcccacattctccatataatccacgcccacacactcatctttaaatt 921
DB 33736 CTACGGTTCCTTCTGGGCCAACATTCCTCATATATCCAGGCACACACTCATTTTAAATAT 33795
QY 922 tagttccacagatctgtactgtgacctttctacactgtagaataacattactactttgtt 981
DB 33796 TAGTTCCAGATCTGTACTGTGACCTTTCTACACTGTGTAATAACATTTACTCATTTTGT 33855
QY 982 caaagaccctctgtgtgctgcttaataatgtagctgactgttttctcctaaaggagtctt 1041
DB 33856 CAAAGACCTTGTGCTGCTTAATATGATGACTGTGTTTTTCTTAAGGAGTGTCT 33915
QY 1042 ggcacagggatctgtaacaggtgggaagcatctcaagatccttccaggggttactact 1101
DB 33916 GSCCCAGGGGATCTGTAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGGTTTACTTT 33975
QY 1102 actagcacacagcatgatacttaacggagtgaattatcttaatacaacatcatctcagtgct 1161
DB 33976 ACTAGCACACAGCATGATCATTTACGGAGTGAATATCTAATAACATCATCTCTCAGTCTC 34035
QY 1162 ttgtcccaactactgaattcattctccactttgtgcccattctcgaacacctcaaaatgct 1221
DB 34036 TTGCCCCATATCGAAATTCATTTCCCACTTTTGCCCCATTCTCAAGACCTCAAAATGTC 34095
QY 1222 attcctaataatcacagattaaactttttttaacctggaagaattcaattgattacat 1281
DB 34096 ATTCCATTAATATCACAGGATTAACCTTTTTTTTTTAAACCTGCAAGAAATTCATGTTACAT 34155
QY 1282 gcagctatgggaatttaattacataattttgtttccagtgcaaaagatgactaagtccttt 1341
DB 34156 GCAGCTATGGGAATTAATATACATATTTTGTTCAGTGCAGAAAGTACTAAGTCTCTT 34215
QY 1342 atccctccctttgttgaattttttccagataaagttaaatgcttagccttgact 1401
DB 34216 ATCCCTCCCTTTGTTGATTTTTTTTCCAGATATAAAGTTAAATGCTTAGCCTGTGACT 34275
QY 1402 gaggctgtatcacg-cacagcctctcccacatccctccagccttatctgtcatcaccaatca 1460
DB 34276 GAGGCTGTATACAGCCACAGCCTCTCCCATCCCTCCAGCCTTATCTGTGTCATCACCATCA 34335
QY 1461 accctcccatnysacctaaacaaatctaaacttgaattccttgaacatgtcaggnact 1520
DB 34336 ACCCTCCCATG-CACCTAAACAAATCTAATCTGTAATTCCTTGAACATGTCAGS-CAT 34393
QY 1521 acatrttctctcctgagagcctcttctctctctctctctctctctctctctctctctctct 1580
DB 34394 ACATATTCTCTCTGCTGAGAGGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 34453
QY 1581 ttttgaataagtgtactacttacttctcgtcaagaagaaggacacataagagattcatcat 1640
DB 34454 TTTTGAATAAGTTGACTATCTTACTTCATGCAAGAGGACACATATGAGATTCATCAT 34513
QY 1641 cacatgagacagcaataactaaaagttaatttgattataagagtttagataataatag 1700
DB 34514 CACATGAGACAGCAATACTAAAAGTGAATTTGATTTATAAGAGTTTAGATAAATATATG 34573
QY 1701 aaatgcagagkccacagaggggaattttatggggcacgttttgaagcctggagatggaag 1760
DB 34574 AAATGCAAGAGCCACAGAGGGGAATGTTTATGGGGCACGCTTTTAAAGCCCTGGGATGGAAG 34633
QY 1761 maaaggcagggaacctcatagtatcttataataataacttcattctctctctctctctctct 1820

```

Db 34634 CAAAGCAGGGAACCTCATAGTATCTTATATAATATACTTCAATTTCTATCTCTCATC 34693
QY 1821 aatattcaacaagcttttccagaattcatgcagtgcaaatcccaaggtaaccttttat 1880
Db 34694 AATATCCACAACGCTTTTACAGAATTTCAGCGAGTCGAATCCCAAGGTAACCTTTAT 34753
QY 1881 ccatttcagtgtgagtcgcttttagaattttggcaaatcatactatgctcattatctcaac 1940
Db 34754 CCATTTTCATGTCAGTCGCTTTAGAAATTTGGCAATCATATGCTGCTACTTATCTCAAC 34813
QY 1941 tttagatgtgtttgtccctgttagttaaattgaaagaatagggcactctgtgagccact 2000
Db 34814 TTTGAGATGTGTTGTCTGTAGTTAATGAAAGAAATAGGCACTCTGTGAGCACT 34873
QY 2001 tttaggttcactctgagcaataagaatttacaaga 2037
Db 34874 TTAGGGTTTCACTCCTGGCAATAAGAAATTTACAAGA 34910

RESULT 4
AL359314
LOCUS      AL359314      173831 bp      DNA      HTG      08-APR-2001
DEFINITION Homo sapiens chromosome 9 clone RP11-108L4, *** SEQUENCING IN
            PROGRESS ***, 3 unordered pieces.
ACCESSION  AL359314
VERSION    AL359314.12 GI:13396560
KEYWORDS   HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 173831)
            Leongamornlert, D.
            Direct Submission
            Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            Requests: clonerequest@sanger.ac.uk
            On Mar 20, 2001 this sequence version replaced gi:13277120.
            ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----- Project Information
            Center project name: ba108L4
            ----- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 173545 bases at least Q40
            Consensus quality: 173599 bases at least Q30
            Consensus quality: 173621 bases at least Q20
            Insert size: 173631; sum-of-contigs
            Insert size: 172123; 10.0% error; agarose-fp
            Quality coverage: 10.35x in Q20 bases; sum-of-contigs Quality
            coverage: 10.50x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            * 1 50595: contig of 50595 bp in length
            * 50596 50695: gap of 100 bp
            * 50696 157636: contig of 106941 bp in length
            * 157637 157736: gap of 100 bp
            * 157737 173831: contig of 16095 bp in length.
            * Location/Qualifiers
            1..173831
FEATURES             source

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-108L4"
/cclone_lib="RPC1-11.1"
1..50595
/note="assembly_fragment:04137
fragment_chain:1"
50696..157636
/note="assembly_fragment:04236
fragment_chain:1"
157737..173831
/note="assembly_fragment:01656"
BASE COUNT 52107 a 36469 c 35018 g 50037 t 200 others
ORIGIN

Query Match      66.8%; Score 1361.4; DB 80; Length 173831;
Best Local Similarity 92.4%; Pred. No. 0;
Matches 1395; Conservative 4; Mismatches 107; Indels 3; Gaps 3;

QY 442 cagagatccctgggagaaatgccggccgcacatcttgggtcgcagatgagcctgcgcctg 501
    |||
Db 156237 CACAGATCCTCGGAGAAATGCCGGCCGCCATCTTGGGTCAATCGATGAGCCTCGCCCTG 156296
    |||

QY 502 tgcctgggtcccgcttgtgaggaagacattagaaaaataattgattgttctcttaagg 561
    |||
Db 156297 TGCCTGGTCCCGCTTGTGAGGAAGACATTAGAAAATGAATGATGTCTTCTTAAGG 156356
    |||

QY 562 atgggcaggaaacagatcctgttgggataatttttgaacgggattacagatttga 621
    |||
Db 156357 ATGGGCAGGAAACAGATCCTGTGTGGGATATTTATTTGAACGGGATTACAGATTTGAAA 156416
    |||

QY 622 tgaagtcaaaagtgcagcattaccgaatgagagaaaaacagagagaaaaatcttgagct 681
    |||
Db 156417 TGAAGTCACAAAGTGAGCATTTACCAATGAGAGGAAACAGAGAGAAAATCTTGATGGCT 156476
    |||

QY 682 tcacaagcatgcacaacaaatggaatactgtgacatgagcagcagcaagctgag 741
    |||
Db 156477 TCACAAGACATGCACAAACAAATGGAATAGTGTGATGACATGAGGCAGCCAGCTGGG 156536
    |||

QY 742 gaggagataaccacgggagcaggggtcaggattctggcctgcctcctaaacttgcgtt 801
    |||
Db 156537 GAGGAGATAACCCAGGGGAGAGGGTCAAGGATTTCTGGCCCTCTGCTGCTAACTGTGCTT 156596
    |||

QY 802 cataaccaatcatctcatatttcaacctcaaacaaagctgttgtaatactgact 861
    |||
Db 156597 CATACCAAAATCATTTATATTTCTAACCTCAAACAAAGCTGTTGTAATATCTGATCT 156656
    |||

QY 862 ctacggttctcttggcccaacattctccatatatccagccacacactatttttaatt 921
    |||
Db 156657 CTACGGTTCTTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCATTTTAAATAT 156716
    |||

QY 922 tagttccagatctgtactgtgaccttttccactgtagaataacattactctttgtt 981
    |||
Db 156717 TAGTTCCCAAGATCTGTACTGTGACCTTTTCTACACTGTAGATAAATTAATTTGTT 156776
    |||

QY 982 caaacacctctgtctgcctaataatgtagctgactgttttcttaagagagttct 1041
    |||
Db 156777 CAAGACCCCTTCTGTGCTGCTTAATATGAGTGCATGTTTCTTCTAAGAGGTGTTCT 156836
    |||

QY 1042 gggccaggggatctgtgaacaggctgggaagcatctcagatctttccagggttatactt 1101
    |||
Db 156837 GGGCCAGGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGCTTATACTT 156896
    |||

QY 1102 actagcacacagcatgatacttacaggagtgaaattatctaatcaacatactcctcagtgctc 1161
    |||
Db 156897 ACTAGCACACAGCATGATCATTTACGGAGTGAATTTATCTAATCAACATCATCTCAGTGTG 156956
    |||

QY 1162 tttagccatactgaattcatttcccaacttttgcctatttgcctcattcctcaagacctcaaatgtc 1221
    |||
Db 156957 TTGGCCCATCTGAAATTCATTTTCCACCTTTTGTGCCCATCTTCTCAAGACCTCAAAATGTC 157016
    |||

```

```
QY 1222 attccattaatcacagaggataaacttttttttttaaccctggaagaattcaattgttacat 1281
|||||
Db 157017 ATTCCATTAAATCACAGGATTAACCTTTTTTTTTTAACTGGAAGATTCATGTTTACAT 157076
QY 1282 cgagctatgggaatttaattacatatatttggttttccagtgcaaaagatgaactaacgttctt 1341
|||||
Db 157077 GCAGCTATGGGAATTAATATACATATTTTGTGTTTCCAGTGCAGAAATGACTAAGTCCTTT 157136
QY 1342 atccctccctctgtttgattttttccagttataaagttaaaatgcttgactgtact 1401
|||||
Db 157137 ATCCCTCCCTCTGTTGTTGATTTTTTTTCCAGTATAAAGTTAAATGCTTAGGCTGTACT 157196
QY 1402 gaggtgtatcacg-cacagcctctcccccattccctccagcccttatctgtcatcacatca 1460
|||||
Db 157197 GAGGCTGTATACAGCCACAGCCTCTCCCATCCCTCCAGCCTTATCTGTCATCACCATCA 157256
QY 1461 accctcccatnysacctaaacaaacttaacttgtaattccttgaacatgctcaggnat 1520
|||||
Db 157257 ACCCTCCCATG-CACCTAAACAAATCTAATCTTAATTCCTTGAACATGTCAGG-CAT 157314
QY 1521 acattttctctctgcctgagagctctctctctctctcttaantctagaatgatgaag 1580
|||||
Db 157315 ACATTATTCCTTCTGCTGAGAGGCTCTCTCTGCTCTTAATCTAGAATGATGTAAG 157374
QY 1581 ttttgaataagttgactatcttacttctcatgcaaaagagacacatatgagattcatcat 1640
|||||
Db 157375 TTTTGAATAAGTTGACTATCTTACTTCTATGCAAGAGGACACATATGAGATTCTATCAT 157434
QY 1641 cacatgagacagcaataactaaaagtgaatttgattataagagtttagataaatatag 1700
|||||
Db 157435 CACATGAGACAGCAATACTAAAGTGTAATTTGATTAATTAAGAGTTTAGATAAATATAG 157494
QY 1701 aaatcaagacacacagaggaagtgttatgggcagcttttgaagcctgggatgtgaag 1760
|||||
Db 157495 AAATGCAAGAGCCACAGAGGGAATGTTTATGGGGCAGCTTTGTAAGCCTGGGATGGAAG 157554
QY 1761 maaagcagggaacctcatagtatcttatataataacttctctctctctctctctctatc 1820
|||||
Db 157555 CAAAGCAGGGAACCTCATAGTATCTTATATAATATCTTCTATCTCTATCTATCATC 157614
QY 1821 aaatcccaacaagcttttcacagaattcatcgagtgcaaatcccccaaggaactcttat 1880
|||||
Db 157615 AATATCCAAACAGCTTTTCACANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 157674
QY 1881 ccatttcagtggtgagcgcttttagaatttgcaaatcatactgctcatctctcaac 1940
|||||
Db 157675 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 157734
QY 1941 tttagatg 1949
|||||
Db 157735 NNACACAAG 157743

RESULT 5
LOCUS AL358573/c
DEFINITION Homo sapiens chromosome 9 clone RP11-133022, *** SEQUENCING IN
PROGRESS ***, 37 unordered pieces.
ACCESSION AL358573
VERSION AL358573.17 GI:13660951
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 267581)
Mclay,K.
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:133988774.
----- Genome Center
```

```
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA133022
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 256599 bases at least Q40
Consensus quality: 259744 bases at least Q30
Consensus quality: 261407 bases at least Q20
Insert size: 263981; sum-of-contigs
Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality
Coverage: 12.23x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 25718: contig of 25718 bp in length
* 25719 25818: gap of 100 bp
* 25819 29567: contig of 3749 bp in length
* 29568 29667: gap of 100 bp
* 29668 40520: contig of 10853 bp in length
* 40521 40620: gap of 100 bp
* 40621 50002: contig of 9382 bp in length
* 50003 50102: gap of 100 bp
* 50103 54942: contig of 4840 bp in length
* 54943 55042: gap of 100 bp
* 55043 59067: contig of 4025 bp in length
* 59068 59167: gap of 100 bp
* 59168 72327: contig of 13160 bp in length
* 72328 72427: gap of 100 bp
* 72428 79396: contig of 6969 bp in length
* 79397 79496: gap of 100 bp
* 79497 82738: contig of 3242 bp in length
* 82739 82838: gap of 100 bp
* 82839 87543: contig of 4705 bp in length
* 87544 87643: gap of 100 bp
* 87644 89847: contig of 2204 bp in length
* 89848 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7323 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length
* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2383 bp in length
* 121870 121969: gap of 100 bp
* 121970 125546: contig of 3577 bp in length
* 125547 125646: gap of 100 bp
* 125647 128990: contig of 3344 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length
* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2198 bp in length
```

```

* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.

```

FEATURES

source

```

1. .267581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/Chromosome="9"
/clone="RP11-133022"
/clone_lib="RPC1-11.1"
1. .25718
/note="assembly_fragment:00645"
fragment_chain:1
clone_end:T7
vector_side:left"
25819..29567
/note="assembly_fragment:01660"
fragment_chain:1"
29668..40520
/note="assembly_fragment:03080"
fragment_chain:1"
40621..50002
/note="assembly_fragment:00771"
fragment_chain:1"
50103..54942
/note="assembly_fragment:02629"
fragment_chain:1"
55043..59067
/note="assembly_fragment:04565"
fragment_chain:1"
59168..72327
/note="assembly_fragment:00223"
fragment_chain:1"
72428..79396
/note="assembly_fragment:03318"
fragment_chain:1"
79497..82738
/note="assembly_fragment:00224"
fragment_chain:2"
82839..87543
/note="assembly_fragment:03050"
fragment_chain:2"
87644..89847
/note="assembly_fragment:01948"
fragment_chain:3"
89948..92376
/note="assembly_fragment:02191"
fragment_chain:3"
92477..99799
/note="assembly_fragment:00010"

```

```

misc_feature
99900..102117
/note="assembly_fragment:00672"
102218..109330
/note="assembly_fragment:01326"
109431..112187
/note="assembly_fragment:01498"
112288..116407
/note="assembly_fragment:01607"
116508..119386
/note="assembly_fragment:01899"
119487..121869
/note="assembly_fragment:02068"
121970..125546
/note="assembly_fragment:02469"
125647..128990
/note="assembly_fragment:02502"
129091..134055
/note="assembly_fragment:02669"
134156..138314
/note="assembly_fragment:02757"
138415..140612
/note="assembly_fragment:02776"
140713..143296
/note="assembly_fragment:03127"
143397..148113
/note="assembly_fragment:03634"
148214..151009
/note="assembly_fragment:03898"
151110..156616
/note="assembly_fragment:04220"
156717..158763
/note="assembly_fragment:04276"
158864..162129
/note="assembly_fragment:04418"
162230..164443
/note="assembly_fragment:04428"

```

```

Query Match      64.7%; Score 1317.8; DB 80; Length 267581;
Best Local Similarity 91.9%; Pred. No. 0;
Matches 1355; Conservative 4; Mismatches 113; Indels 3; Gaps 3;

```

```

QY 442 cagagatccctgggagaaatgccggcgcccatcttgggtcactgatgagcctcgccctg 501
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165900 CACAGATCCCTGGGAGAAATGCCGGCGCCCATCTTTGGGTCAATCGATAGCCTCGCCCTG 165841

QY 502 tcctgtgtccctgtgtgaggaagacattagaaataaattgattgttctcctaaag 561
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165840 TGCTGTGTCCTGTGTGAGGGAAGGACATTAGAAATGAATGTGTGTCTTTAAGG 165781

QY 562 atgggcaggaacacagatcctgtgtgtggatattatttgaacgggattacagatttga 621
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165780 ATGGGCAGGAAACACAGATCCTGTGTGGATATTTATTGAACGGGATTACAGATTTGAAA 165721

QY 622 tgaagtacacaaagtggagcattaccgaatgaggaagaaacagagagaaatcttgatggct 681
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165720 TGAAGTCACAAAAGTGAGCATTACCAATGAGAGAAACACAGAGAAAATCTTTGATGGCT 165661

QY 682 tcacaagacatcaacaaacaaatggaatactgtgatacactgagcagcagcaagctgg 741
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165660 TCACAAGACATGCAACAAACAAATGGAATACGTGTGATGACATGAGCAGCAAGCTGGG 165601

QY 742 gaggagataaccacggggcagaggggtcaggattcttggcctgctgcctaaactgtcgctt 801
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165600 GAGGAGATAACACACGGGCGAGAGGGTCAAGATTCTGCGCCTGCTGCTTAACTGTGGTT 165541

QY 802 cataaccaaatactttcatatttcttaacctcaaaacaaagctgttgcataatcctgatct 861
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165540 CATAACCAAATCATTTTCATATTTCTTAACCTCAAAACAAAGCTGTTGTATATCTGATCT 165481

QY 862 ctacggttctcttgggcccaacattctccatatatccagccacacactcattttatatt 921
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165480 CTACGGTTCCTTCTGGGGCCCAACATTCTCCATATATATCCAGCCCACTCAATTTTAAATAT 165421

```



```
QY 922 tagtccacagatctgactgagacctttctacactgtagaataacattactatttgg 981
|||||
Db 165420 TAGTTCCCGAGATCTGACTGTGACCTTCTACACTGTAGATAACATTACTATTTGGT 165361
|||||
QY 982 caaagaaccttgctgctgcttaataatgtagctgactggttttccctaaaggagtgctc 1041
|||||
Db 165360 CAAGAACCTTCGTGCTGCTGAATATGTAGCTGACTGTTTTCCTAAAGGAGGTGCT 165301
|||||
QY 1042 gcccaggggactctgaaacagctgggaagcatctcaagatcttccaggggttactt 1101
|||||
Db 165300 GGCCAGGGGATCTGTGAACAGCTGGGAAGCATCTCAAGATCTTCCAGGGTATACTT 165241
|||||
QY 1102 actagcacacagcatgacattaccggagtggaattatctaatcaacatcatctcagtgctc 1161
|||||
Db 165240 ACTAGCACACAGCATGATCATTTACGGAGTGAATTAATCAATCAACATCATCTCAGTGTC 165181
|||||
QY 1162 ttggccactactgaaattcatttcccaacttttggccatttctcaagacctcaaaatgctc 1221
|||||
Db 165180 TTGGCCCATACTGAAATTAATTTCCACCTTTTGTGCCATTTCTCAAGACCTCAAAATGTC 165121
|||||
QY 1222 attccattaatcacagattaaacttttttttaacctgggaagaattcaatgtttacat 1281
|||||
Db 165120 ATTCCATTAAATACAGGATTAATCTTTTTCAGTATAAAGTTAAATGCTTAGCTGTACT 165061
|||||
QY 1282 cgagctatgggaatttaattacatatattttgtttccagtgcaagatgactaagtccttt 1341
|||||
Db 165060 GCAGCTATGGGAATTAATTAATATATTTGTTTCCAGTGCAAGATGACTAAGTCCITT 165001
|||||
QY 1342 atccctcccttggttgatttttttccagataaagttaaaatgcttagccttgtaact 1401
|||||
Db 165000 ATCCCTCCCTTGTGTGATTTTTCAGTATAAAGTTAAATGCTTAGCTGTACT 164941
|||||
QY 1402 gaggctgtatacag-cacagcctctcccatccctccagccttactctgtcatcacatca 1460
|||||
Db 164940 GAGGCTGTATACAGCAGCAGCTCTCCCATCCCTCCAGCTTACTGTGTCATCACCATCA 164881
|||||
QY 1461 accctcccatnysacctaaacaaatcctaactgttaattccttgaacatgtcaggncat 1520
|||||
Db 164880 ACCCTCCCATG-CACCTAAACAAATCTAATCTGTAATTCCTTGAACATGTCAGG-CAT 164823
|||||
QY 1521 acattttcctctgcctgagagctcttctctctctctctctctctctctctctctctctct 1580
|||||
Db 164822 ACATTATTCTCTGCTGCTGAGAGCTCTTCCCTGTCTCTTAAATCTAGAATGATGTAAG 164763
|||||
QY 1581 ttttgaataagtgactatttacttctcatgcaagaaggacacatatgagattcatcat 1640
|||||
Db 164762 TTTTGAATAAGTTGACTATCTTACTTTCATGCAAGAGGACACATATGAGATTCTATCAT 164703
|||||
QY 1641 cacatgacagcaataactaaaagttaattttgattataagagttttagataaataatg 1700
|||||
Db 164702 CACATGAGACAGCAATACTAAAGTGTAATTTGATTATAACAGTTTACATAAATATATG 164643
|||||
QY 1701 aaatgcagagkccacagaggggaatgtttatgggggacgttttgaagcctggagatgtaag 1760
|||||
Db 164642 AAATGCAGAGGACAGAGGGGAATGTTTATGGGGACGCTTTGTAAGCCTGGGATGTGAAG 164583
|||||
QY 1761 maaagcaggggaacctcatgattcttataataataatccttctctctctctctctctct 1820
|||||
Db 164582 CAAGGAGGAGCACTCATGATCTTATATATATATATATATATATATATATATATATAT 164523
|||||
QY 1821 aataatcaacaagcgtttcacagaattcatgagtgcaaatcccccaaggtaacctttat 1880
|||||
Db 164522 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 164463
|||||
QY 1881 ccatttcatggtgagtgoccttagaatttggca 1915
|||||
Db 164462 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 164428
|||||
RESULT 6
AX018075 359 bp DNA PAT 07-SEP-2000
LOCUS Sequence 23 from Patent W09946374.
DEFINITION
```

```
AX018075
VERSION AX018075.1 GI:10042526
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 359)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarsky,C.
TITLE Human nucleic acid sequences from prostate tumour tissue
JOURNAL Patent: WO 9946374-A 23 16-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
source 1..359
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 121 a 75 c 94 g 69 t
ORIGIN
Query Match 11.6%; Score 235.4; DB 9; Length 359;
Best Local Similarity 99.6%; Pred. No. 6.8e-48;
Matches 236; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 259 cagggtgagaaataagaaggctgctgactttaccatctgaggccacacatctgctgaa 318
|||||
Db 123 CAGAGTGTGAGAAATAGAAGGCTGCTGACTTTTACCATCTGAGGCCACACATCTGCTGAA 182
|||||
QY 319 atggagataaattaacatcactagaacagcaagatgacaataatgtctaaagttagtgc 378
|||||
Db 183 ATGGAGATAATTAACTACTAGAACACAGCAAGATGACAAATATATGCTTAAGTAGTGAC 242
|||||
QY 379 atgttttgcacatttccagccctttaaataatccacacacaggaagacacaaaggaa 438
|||||
Db 243 ATGTTTTTCACATTTCAGCCCTTTAAATATCCACACACACAGGAAGCACAAAAGGAA 302
|||||
QY 439 gcacagagatccctgggagaaatcccgccgcatcttgggtcatcgatgagctc 495
|||||
Db 303 GCACAGAGATCCCTGGGAGAAATGCCCGCCCATCTTGGGTCTCATCGATGAGCCCTC 359
|||||
RESULT 7
AL161625
LOCUS AL161625 143675 bp DNA HTG 20-JAN-2001
DEFINITION Homo sapiens chromosome 9 clone RP11-146P9, *** SEQUENCING IN
PROGRESS ***, 10 unordered pieces.
ACCESSION AL161625
VERSION AL161625.6 GI:9863607
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 143675)
AUTHORS Plumb,B.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:8894260.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bal46P9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
```

Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 138647 bases at least Q40
Consensus quality: 140619 bases at least Q30
Consensus quality: 141706 bases at least Q20
Insert size: 142775; sum-of-contigs
Insert size: 147523; 5.9% error; agarose-fp
Quality coverage: 4.29x in Q20 bases; sum-of-contigs Quality
coverage: 4.22x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 12163: contig of 12163 bp in length
* 12164 12263: gap of 100 bp
* 12264 25269: contig of 13006 bp in length
* 25270 25369: gap of 100 bp
* 25370 57709: contig of 32340 bp in length
* 57710 57809: gap of 100 bp
* 57810 66792: contig of 8983 bp in length
* 66793 66892: gap of 100 bp
* 66893 102772: contig of 35880 bp in length
* 102773 102872: gap of 100 bp
* 102873 106863: contig of 3991 bp in length
* 106864 106963: gap of 100 bp
* 106964 129804: contig of 22841 bp in length
* 129805 129904: gap of 100 bp
* 129905 132223: contig of 2319 bp in length
* 132224 132323: gap of 100 bp
* 132324 135764: contig of 3441 bp in length
* 135765 135864: gap of 100 bp
* 135865 143675: contig of 7811 bp in length.

FEATURES

source

1. .143675
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone_lib="RPC1-11.1"
/clone_end:77
/note="assembly_fragment:00440
vector_side:left"
12264..25269
/note="assembly_fragment:00434
fragment_chain:1"
25370..57709
/note="assembly_fragment:00526
fragment_chain:1"
57810..66792
/note="assembly_fragment:00661
fragment_chain:1"
66893..102772
/note="assembly_fragment:00323"
102873..106863
/note="assembly_fragment:00860"
106964..129804
/note="assembly_fragment:00978"
129905..132223
/note="assembly_fragment:01068"
132324..135764
/note="assembly_fragment:01424"
135865..143675
/note="assembly_fragment:01413
clone_end:sp6
vector_side:right"

BASE COUNT 42329 a 30042 c 28964 g 41431 t 909 others
ORIGIN

Query Match 4.9%; Score 99; DB 79; Length 143675;
Best Local Similarity 100.0%; Pred. No. 8.9e-14;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agaagctgcattcagaaaaaacagagggagattgtgtgctgcagccgagggagaccag 60
Db 136065 AGAAGCTGCATCAGAAAAACAGAGGGAGATTGTGTGCTGCAGCCGAGGAGACCAG 136124
QY 61 gaagatctcatgggtgggaagaccctgatgatacagagg 99
Db 136125 GAAGATCTCATGGTGGGAAGGACCTGATGATACAGAGG 136163

RESULT 8

AL358573

LOCUS AL358573 267581 bp DNA HTG 15-APR-2001
DEFINITION Homo sapiens chromosome 9 clone RP11-133022, *** SEQUENCING IN
PROGRESS ***, 37 unordered pieces.

ACCESSION AL358573

VERSION AL358573.17 GI:13660951

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 267581)

AUTHORS Mclay, K.

TITLE Direct Submission

JOURNAL Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13398774.

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BAL33022

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 256699 bases at least Q40

Consensus quality: 259744 bases at least Q30

Consensus quality: 261407 bases at least Q20

Insert size: 263981; sum-of-contigs

Insert size: 135491; 19.3% error; agarose-fp

Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality

coverage: 12.23x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 37 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 25718: contig of 25718 bp in length

* 25719 25818: gap of 100 bp

* 25819 29567: contig of 3749 bp in length

* 29568 29667: gap of 100 bp

* 29668 40520: contig of 10853 bp in length

* 40521 40620: gap of 100 bp

* 40621 50002: contig of 9382 bp in length

* 50003 50102: gap of 100 bp

* 50103 54942: contig of 4840 bp in length

* 54943 55042: gap of 100 bp

* 55043 59067: contig of 4025 bp in length

* 59068 59167: gap of 100 bp

* 59168 72327: contig of 13160 bp in length

* 72328 72327: contig of 13160 bp in length

```

* 72328 72427: gap of 100 bp
* 72428 79396: contig of 6969 bp in length
* 79397 79496: gap of 100 bp
* 79497 82738: contig of 3242 bp in length
* 82739 82838: gap of 100 bp
* 82839 87543: contig of 4705 bp in length
* 87544 87643: gap of 100 bp
* 87644 89847: contig of 2204 bp in length
* 89848 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7323 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length
* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2383 bp in length
* 121870 121969: gap of 100 bp
* 121970 125546: contig of 3577 bp in length
* 125547 125646: gap of 100 bp
* 125647 128990: contig of 3344 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length
* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2198 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.
FEATURES
    source
        1..267581
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone="RP11-133022"
            /clone.lib="RPC1-11.1"
        1..25718
            /note="assembly_fragment:00645
            fragment_chain:1
            clone_end:T7
            vector_side:left"
        25819..29567

```

```

misc_feature
    /note="assembly_fragment:01660
    fragment_chain:1"
    29668..40520
    /note="assembly_fragment:03080
    fragment_chain:1"
    40621..50002
    /note="assembly_fragment:00771
    fragment_chain:1"
    50103..54942
    /note="assembly_fragment:02629
    fragment_chain:1"
    55043..59067
    /note="assembly_fragment:04565
    fragment_chain:1"
    59168..72327
    /note="assembly_fragment:00223
    fragment_chain:1"
    72428..79396
    /note="assembly_fragment:03318
    fragment_chain:1"
    79497..82738
    /note="assembly_fragment:00224
    fragment_chain:2"
    82839..87543
    /note="assembly_fragment:03050
    fragment_chain:2"
    87644..89847
    /note="assembly_fragment:01948
    fragment_chain:3"
    89948..92376
    /note="assembly_fragment:02191
    fragment_chain:3"
    92477..99799
    /note="assembly_fragment:00010"
    99900..102117
    /note="assembly_fragment:00672"
    102218..109330
    /note="assembly_fragment:01326"
    109431..112187
    /note="assembly_fragment:01498"
    112288..116407
    /note="assembly_fragment:01607"
    116508..119386
    /note="assembly_fragment:01899"
    119487..121869
    /note="assembly_fragment:02068"
    121970..125546
    /note="assembly_fragment:02469"
    125647..128990
    /note="assembly_fragment:02502"
    129091..134055
    /note="assembly_fragment:02669"
    134156..138314
    /note="assembly_fragment:02757"
    138415..140612
    /note="assembly_fragment:02776"
    140713..143296
    /note="assembly_fragment:03127"
    143397..148113
    /note="assembly_fragment:03634"
    148214..151009
    /note="assembly_fragment:03898"
    151110..156616
    /note="assembly_fragment:04220"
    156717..158763
    /note="assembly_fragment:04276"
    158864..162129
    /note="assembly_fragment:04418"
    162230..164443
    /note="assembly_fragment:04428"

```

Query Match 4.9%; Score 99; DB 80; Length 267581;
 Best Local Similarity 100.0%; Pred. No. 9.4e-14;

```

Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agaagctgcacagaaacagagggagattgtgtgctgcagccgaggagaccag 60
|||||
Db 253275 AGAAGCTGCATCAGAAAACAGAGGGAGATTGTGTGGCTGCACCCGAGGAGACCAG 253334
|||||
QY 61 gaagatctgcattgtgtgggaagacctgatgatcacagag 99
|||||
Db 253335 GAAGATCTGCATGTTGGGAAGGACCTGATGATACAGAGG 253373
|||||

RESULT 9
AF279290 580 bp DNA PRI 04-DEC-2000
LOCUS Homo sapiens prostate-cancer-specific DD3 protein gene, promoter
DEFINITION AF279290
ACCESSION AF279290
VERSION AF279290.1 GI:11528086
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 580)
AUTHORS Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and
Bussemakers,M.J.G.
TITLE Isolation and Characterization of the Promoter of the Human
Prostate Cancer-specific DD3 Gene
JOURNAL J. Biol. Chem. 275 (48), 37496-37503 (2000)
PUBMED 10982808
REFERENCE 2 (bases 1 to 580)
AUTHORS Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and
Bussemakers,M.J.G.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-2000) Urology Research Laboratory, University
Medical Center Nijmegen, Geert Grooteplein Zuid 10, Nijmegen 6525
GA, The Netherlands

FEATURES
Source
1..580
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q21-q22"
1..460
/promoter 260..288
/protein_bind 260..288
/mRNA 461..580
/product="prostate-cancer-specific DD3 protein"
/exon 461..580
BASE COUNT 190 a 97 c 153 g 140 t
ORIGIN

Query Match 4.8%; Score 98; DB 89; Length 580;
Best Local Similarity 100.0%; Pred. No. 9.5e-14;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agaagctgcacagaaacagagggagattgtgtgctgcagccgaggagaccag 60
|||||
Db 483 AGAAGCTGCATCAGAAAACAGAGGGAGATTGTGTGGCTGCACCCGAGGAGACCAG 542
|||||
QY 61 gaagatctgcattgtgtgggaagacctgatgatcacagag 98
|||||
Db 543 GAAGATCTGCATGTTGGGAAGGACCTGATGATACAGAG 580
|||||

RESULT 10
AC009556/c
LOCUS Homo sapiens clone RP11-57C21, WORKING DRAFT SEQUENCE, 26 unordered
DEFINITION pieces
ACCESSION AC009556

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
COMMENT

```

```

AC009556.4 GI:7107752
HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172298)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome, clone RP11-57C21
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 172298)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
Funke,B., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Feb 28, 2000 this sequence version replaced gi:6479158.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1672
Center clone name: 57_C21
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-primer-amersham; 5% of reads
Chemistry: Dye-terminator Big Dye; 95% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 136385 bases at least Q40
Consensus quality: 156343 bases at least Q30
Consensus quality: 165534 bases at least Q20
Insert size: 165000; agarose-fp
Insert size: 169798; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 base.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1016: contig of 1016 bp in length
* 1017 1116: gap of 100 bp
* 1117 2136: contig of 1020 bp in length
* 2137 2236: gap of 100 bp
* 2237 3571: contig of 1335 bp in length
* 3572 3671: gap of 100 bp
* 3672 5272: contig of 1601 bp in length
* 5273 5372: gap of 100 bp
* 5373 6381: contig of 1009 bp in length
* 6382 6481: gap of 100 bp
* 6482 8508: contig of 2027 bp in length
* 8509 8608: gap of 100 bp

```

```
* 8609 9886: contig of 1278 bp in length
* 9887 9986: gap of 100 bp
* 11943 11943: contig of 1957 bp in length
* 11944 12043: gap of 100 bp
* 12044 15112: contig of 3069 bp in length
* 15113 15212: gap of 100 bp
* 15213 19323: contig of 4111 bp in length
* 19324 19423: gap of 100 bp
* 19424 24065: contig of 4642 bp in length
* 24066 24165: gap of 100 bp
* 24166 31658: contig of 7493 bp in length
* 31659 31758: gap of 100 bp
* 31759 38737: contig of 6979 bp in length
* 38738 38837: gap of 100 bp
* 38838 46131: contig of 7294 bp in length
* 46132 46231: gap of 100 bp
* 46232 52344: contig of 6113 bp in length
* 52345 52444: gap of 100 bp
* 52445 59763: contig of 7319 bp in length
* 59764 59863: gap of 100 bp
* 59864 66816: contig of 6953 bp in length
* 66817 66916: gap of 100 bp
* 66917 74010: contig of 7094 bp in length
* 74011 74110: gap of 100 bp
* 74111 80599: contig of 6489 bp in length
* 80600 80699: gap of 100 bp
* 80700 89507: contig of 8808 bp in length
* 89508 89607: gap of 100 bp
* 89608 100337: contig of 10730 bp in length
* 100338 100437: gap of 100 bp
* 100438 114298: contig of 13861 bp in length
* 114299 114398: gap of 100 bp
* 114399 128530: contig of 14132 bp in length
* 128531 128630: gap of 100 bp
* 128631 142917: contig of 14287 bp in length
* 142918 143017: gap of 100 bp
* 143018 156243: contig of 13226 bp in length
* 156244 156344: gap of 100 bp
* 156344 172298: contig of 15955 bp in length.
```

FEATURES

```
source
1..172298
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-57C21"
/clone_lib="RP11-11 Human Male BAC"
1..1016
/note="assembly_fragment"
1117..2136
/note="assembly_fragment"
2237..3571
/note="assembly_fragment"
3672..5272
/note="assembly_fragment"
5373..6381
/note="assembly_fragment"
6482..8508
/note="assembly_fragment"
8609..9886
/note="assembly_fragment"
9987..11943
/note="assembly_fragment"
12044..15112
/note="assembly_fragment"
15213..19323
/note="assembly_fragment"
19424..24065
/note="assembly_fragment"
24166..31658
/note="assembly_fragment"
31759..38737
/note="assembly_fragment"
38838..46131
/note="assembly_fragment"

misc_feature
46232..52344
/note="assembly_fragment"
52445..59763
/note="assembly_fragment"
59864..66816
/note="assembly_fragment"
66917..74010
/note="assembly_fragment"
74011..80599
/note="assembly_fragment"
80700..89507
/note="assembly_fragment"
89608..100337
/note="assembly_fragment"
100438..114298
/note="assembly_fragment"
114399..128530
/note="assembly_fragment"
128631..142917
/note="assembly_fragment"
143018..156243
/note="assembly_fragment"
156344..172298
/note="assembly_fragment"

BASE COUNT 51330 a 34914 c 34487 g 49065 t 2502 others
ORIGIN

Query Match 4.8%; Score 97.4; DB 61; Length 172298;
Best Local Similarity 99.0%; Pred. No. 2.3e-13;
Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agaagctgcgcacagaaacagaggagagattgttgctgcagccgagggagaccag 60
|||||
Db 16539 AGAGCTGCGCATCAGAAAACAGAGGGGAGATTGTGTGGCTGCAGCCGAGGAGACCAG 16480

Qy 61 gaagatcgcgcagtgtgggaaggacctgatcacagag 99
|||||
Db 16479 GAAGATTGTCATGTTGGGAGGACCTGATACAGAG 16441

RESULT 11
166494
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
source
1..7218
/organism="unknown"

BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 2.9%; Score 59.8; DB 10; Length 7218;
Best Local Similarity 5.1%; Pred. No. 0.00037;
Matches 20; Conservative 219; Mismatches 155; Indels 0; Gaps 0;

Qy 1178 ttcatttccacttttgcctattctcaagacctcaaatgtcattccattatcac 1237
::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1076 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1135
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:50:00 ; Search time 46.06 Seconds
(without alignments)
146.495 Million cell updates/sec

Title: US-09-402-713A-2

Perfect score: 268

Sequence: 1 MFLHSSPPKYPHTQEAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phase:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.unclassified:*
- 13: sp.vertebrate:*
- 14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	60.5	22.6	1236	5 Q9NKF9	Q9nkf9 drosophila
2	60.5	22.6	1238	5 Q9VJW9	Q9vjw9 drosophila
3	60.5	22.6	1239	5 Q94902	Q94902 drosophila
4	60	22.4	187	2 Q9JUF3	Q9juf3 neisseria m
5	60	22.4	218	2 Q9JUF5	Q9juf5 neisseria m
6	59	22.0	1114	11 Q9JL73	Q9j173 mus musculus
7	56.5	21.1	1061	2 P73998	P73998 synechocyst
8	56.5	21.1	6396	2 Q9KID7	Q9kid7 streptomyces
9	56	20.9	338	2 Q9RVU4	Q9rvu4 deinococcus
10	56	20.9	433	5 Q9V7N7	Q9v7n7 drosophila
11	56	20.9	558	5 Q9NFP1	Q9nfp1 drosophila
12	55.5	20.7	698	2 Q9FLJ5	Q9flj5 pseudomonas
13	55	20.5	168	14 Q9E138	Q9e138 bovine vira
14	55	20.5	196	5 Q9VKF5	Q9vkf5 drosophila
15	55	20.5	469	5 Q9GRT5	Q9grt5 leishmania
16	55	20.5	1032	11 Q61989	Q61989 mus musculus
17	55	20.5	1935	5 Q9VQW0	Q9vqw0 drosophila
18	54.5	20.3	116	14 Q92886	Q92886 human immun
19	54.5	20.3	177	6 Q9XS95	Q9xs95 oryctolagus

20	54.5	20.3	244	2 Q9HYW1	Q9hyw1 pseudomonas
21	54.5	20.3	305	10 Q9MIC3	Q9mic3 arabisdopsis
22	54.5	20.3	461	11 Q9RIY8	Q9riy8 mus musculus
23	54	20.1	204	3 Q9P3L8	Q9p3l8 neurospora
24	54	20.1	257	2 Q9Z8E3	Q9z8e3 chlamydia p
25	54	20.1	450	10 Q9FMN6	Q9fmn6 arabisdopsis
26	54	20.1	633	2 Q9HTU6	Q9htu6 pseudomonas
27	54	20.1	784	4 Q00146	Q00146 homo sapien
28	53.5	20.0	250	8 Q33122	Q33122 schismocarp
29	53.5	20.0	348	8 Q35379	Q35379 paracyprich
30	53.5	20.0	408	4 Q9UGL4	Q9ugl4 homo sapien
31	53.5	20.0	421	8 Q9TIC2	Q9tic2 balbiana i
32	53.5	20.0	443	8 Q32878	Q32878 palmaria pa
33	53.5	20.0	449	2 Q9Z9E0	Q9z9e0 chlamydia p
34	53.5	20.0	783	4 Q9NXD0	Q9nxd0 homo sapien
35	53	19.8	671	5 Q9W365	Q9w365 drosophila
36	53	19.8	1161	11 Q9QRE7	Q9qre7 rattus norv
37	53	19.8	2205	5 Q9NGQ2	Q9ngq2 dictyosteli
38	52.5	19.6	216	2 Q9RVK0	Q9rvk0 deinococcus
39	52.5	19.6	357	2 Q9I720	Q9i720 pseudomonas
40	52.5	19.6	360	4 Q9H9N9	Q9h9n9 homo sapien
41	52.5	19.6	384	8 Q32401	Q32401 halosaccion
42	52.5	19.6	392	8 Q31855	Q31855 audouinella
43	52.5	19.6	424	8 Q9MW15	Q9mw15 ballia call
44	52.5	19.6	488	8 Q9THF8	Q9thf8 palmaria pa
45	52.5	19.6	493	4 Q9UF35	Q9uf35 homo sapien

ALIGNMENTS

RESULT 1

ID	Q9NKF9	PRELIMINARY;	PRT;	1236 AA.
AC	Q9NKF9;			
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)			
DT	01-MAR-2001 (TREMBlrel. 16, Last annotation update)			
DE	KUZ PROTEIN.			
CN	KUZ OR BG:DS07660.3 OR CG7147.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=Y, CN BW SP;			
RX	MEDLINE=99403001; PubMed=10471707;			
RA	Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T., Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D., Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A., Palazzi M., Reese M.G., Spradling A., Tsang C., Wan K., Whitelaw K., Ceiniker S., Rubin G.M.;			
RT	"An exploration of the sequence of a 2.9-Mb region of the genome of Drosophila melanogaster: the Adh region."			
RL	Genetics 153:179-219(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Y, CN BW SP;			
RA	Celniker S.E., Aqbayani A., Arcaina T.T., Baxter E., Blazej R.G., Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M., Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A., Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B., Lewis S., Li P., Lomtan M.A., Mazda P., Moshrefi A.R., Moshrefi M., Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A., Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R., Zieran L.L., Rubin G.M.;			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AE003407; AAF4800.1; -;			
DR	FlyBase; FBgn0015954; kuz.			
DR	InterPro; IPR000130; -;			
DR	InterPro; IPR001590; -;			

DR	EMBL; AE003640; AAF53318.1; -.
DR	HSSP; P17494; 1KST.
DR	FlyBase; FBgn0015954; kuz.
DR	InterPro; IPR000130; -.
DR	InterPro; IPR0001590; -.
DR	InterPro; IPR001762; -.
DR	PROSITE; PS50215; ADAM_MEPRO; 1.
DR	PROSITE; PS50214; DISINTEGRIN_2; 1.
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR	SMART; SM00050; DISIN; 1.
SO	SEQUENCE 1238 AA; 136407 MW; 96D924B6188AC472 CRC64;
Query Match 22.6%; Score 60.5; DB 5; Length 1238;	
Best Local Similarity 47.2%; Pred. No. 19;	
Matches 17; Conservative Mismatches 10; Indels 5; Gaps 2;	
QY	9 FKYPHTQEAQKEAQRSLGEM----PGRHLGSSMSLA 40
DB	340 KYPH-QKYTKEANFAEGAFYDPSTGRRLGSSANVA 374 : : :
RESULT 3	
Q94902	PRELIMINARY; PRT; 1239 AA.
ID	Q94902
AC	Q94902;
DT	01-FEB-1997 (TREMBlrel. 02, Created)
DT	01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT	01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE	KUZBANIAN.
OS	KUZ OR BG:DS07660.3 OR CG7147.
GN	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=96355647; PubMed=8703057;
RA	Rooke J., Pan D., Xu T., Rubin G.M.;
RT	"Kuz, a conserved metalloprotease-disintegrin protein with two roles
RL	in Drosophila neurogenesis."
RL	Science 273:1227-1231(1996);
DR	EMBL; U60591; AAC47275.1; -.
DR	HSSP; P17494; 1KST.
DR	MEROPS; M12.211; -.
DR	FlyBase; FBgn0015954; kuz.
DR	InterPro; IPR000130; -.
DR	InterPro; IPR001590; -.
DR	InterPro; IPR001762; -.
DR	PROSITE; PS50215; ADAM_MEPRO; 1.
DR	PROSITE; PS50214; DISINTEGRIN_2; 1.
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR	SMART; SM00050; DISIN; 1.
KW	Neurogenesis.
SO	SEQUENCE 1239 AA; 136522 MW; 7FCB514A1352577A CRC64;
Query Match 22.6%; Score 60.5; DB 5; Length 1239;	
Best Local Similarity 47.2%; Pred. No. 19;	
Matches 17; Conservative Mismatches 10; Indels 5; Gaps 2;	
QY	9 FKYPHTQEAQKEAQRSLGEM----PGRHLGSSMSLA 40
DB	341 KYPH-QKYTKEANFAEGAFYDPSTGRRLGSSANVA 375 : : :
RESULT 4	
Q9JUF3	PRELIMINARY; PRT; 187 AA.
ID	Q9JUF3
AC	Q9JUF3;
DT	01-OCT-2000 (TREMBlrel. 15, Created)
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)


```
DR Pfam: PF00873; ACR_tran; 1.
DR PRINTS: PR00702; ACRIFLAVINRP.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
SQ SEQUENCE 1061 AA; 114935 MW; 480D9B60FC09F8E2 CRC64;

Query Match 21.1%; Score 56.5; DB 2; Length 1061;
Best Local Similarity 31.0%; Pred. No. 55;
Matches 13; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 5 ISSPFYKPTQEAQKQSLGEMPCRHGSSSLALCLVPL 46
   || | | | | | | | | | | | | | | | | | | |
Db 418 ISRFIKVKH-EDPQEAQAQAMGELTGAVIATSLVLMVAFIPV 458

RESULT 8
Q9KID7 PRELIMINARY; PRT; 6396 AA.
ID Q9KID7;
AC Q9KID7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE FKBA.
GN FKBA.
OS Streptomyces hygroscopicus var. ascomyceticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=132248;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20323220; PubMed=10863099;
RA Wu K., Chung L., Revill W.P., Katz L., Reeves C.D.;
RT "The FK520 gene cluster of streptomyces hygroscopicus var.
RT ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual
RT polyketide extender units.";
RL Gene 251:81-90(2000).
DR EMBL; AF235504; AAF86396.1; -.
DR InterPro; IPR000255; -.
DR InterPro; IPR000794; -.
DR InterPro; IPR001066; -.
DR InterPro; IPR001227; -.
DR InterPro; IPR001899; -.
DR InterPro; IPR002085; -.
DR Pfam; PF00107; adh_zinc; 2.
DR Pfam; PF00109; ketoacyl-synt; 4.
DR Pfam; PF00550; pp-binding; 4.
DR Pfam; PF00698; ACYL_transf; 4.
DR PROSITE; PS00075; ACP_DOMAIN; 4.
DR PROSITE; PS00506; B_KETOACYL-SYNTHASE; 4.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_3.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_3.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 6396 AA; 669881 MW; FF13BC2431D2E805 CRC64;

Query Match 21.1%; Score 56.5; DB 2; Length 6396;
Best Local Similarity 35.6%; Pred. No. 3.9e+02;
Matches 16; Conservative 6; Mismatches 14; Indels 9; Gaps 2;

QY 12 PHT-----QEAQKQSLG---EMPGHGLSSMSLAICLVPLV 47
   ||: || || || || || || || || || || || || || || ||
Db 653 PHSVLSDGEAEVLEAARGLGHHRLPTRHAGHSRMPQLVAPLL 697

RESULT 9
Q9RVU4 PRELIMINARY; PRT; 338 AA.
ID Q9RVU4;
AC Q9RVU4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ENDONUCLEASE III; PUTATIVE.
```

```
GN DR0928.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RI;
RC MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RI.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001945; AAF10505.1; -.
DR HSSP; P20625; 2ABK.
DR TIGR; DR0928; -.
DR InterPro; IPR003265; -.
DR Pfam; PF00730; HHH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
KW Endonuclease.
SQ SEQUENCE 338 AA; 38260 MW; 77407CB6E0826BFA CRC64;

Query Match 20.9%; Score 56; DB 2; Length 338;
Best Local Similarity 38.1%; Pred. No. 19;
Matches 8; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 10 KYPHTQEAQKQSLGEMPG 30
   ::||| | ::||| ::||
Db 207 EFPHPGEGQAQALALPG 227

RESULT 10
Q9V7N7 PRELIMINARY; PRT; 433 AA.
ID Q9V7N7;
AC Q9V7N7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE VHA44 PROTEIN.
GN VHA44 OR CG8048.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
```

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.H., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AEO03807; AAF58013.1; -;
 DR FlyBase: FBgn020611; Vha44;
 SQ SEQUENCE 433 AA; 46620 MW; FD29F44E6F317C46 CRC64;

Query Match 20.9%; Score 56; DB 5; Length 433;
 Best Local Similarity 40.7%; Pred. No. 24;
 Matches 11; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 4 HTSSPFKYPHTQEAQKQSRISGEMPG 30
 ID ID Q9NFP1 PRELIMINARY; PRT; 558 AA.
 AC Q9NFP1;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE VANIN-LIKE PROTEIN.
 GN VANIN-LIKE OR CG33648.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Galland F.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99432008; PubMed=10501839;
 RA Granjeaud S., Naquet P., Galland F.;
 RT "An EST description of the new Vanin gene family conserved from fly to human.";
 RL Immunogenetics 49:964-972(1999).
 DR EMBL: AJ276261; CAB77020.1; -;
 DR FlyBase: FBgn0040069; vanin-like.
 DR InterPro: IPR003010; -;
 SQ SEQUENCE 558 AA; 62341 MW; 7920621923BB7779 CRC64;

RESULT 11
 Q9NFP1

ID ID Q9NFP1 PRELIMINARY; PRT; 558 AA.
 AC Q9NFP1;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE VANIN-LIKE PROTEIN.
 GN VANIN-LIKE OR CG33648.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Galland F.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99432008; PubMed=10501839;
 RA Granjeaud S., Naquet P., Galland F.;
 RT "An EST description of the new Vanin gene family conserved from fly to human.";
 RL Immunogenetics 49:964-972(1999).
 DR EMBL: AJ276261; CAB77020.1; -;
 DR FlyBase: FBgn0040069; vanin-like.
 DR InterPro: IPR003010; -;
 SQ SEQUENCE 558 AA; 62341 MW; 7920621923BB7779 CRC64;

Query Match 20.9%; Score 56; DB 5; Length 558;
 Best Local Similarity 33.3%; Pred. No. 32;
 Matches 12; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 3 LHSSPFKYPHTQEAQKQSRISGEMPG 38
 ID ID Q9VKF5 PRELIMINARY; PRT; 196 AA.
 AC Q9VKF5;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)

RESULT 12

Q9FLJ5 PRELIMINARY; PRT; 698 AA.
 ID ID Q9FLJ5
 AC Q9FLJ5;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE PUTATIVE HRPO.
 GN PUTATIVE HRPO.
 OS Pseudomonas glumae.
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
 OC Burkholderia.
 OX NCBI_TaxID=337;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GM4;
 RA Takikawa Y., Eo S., Adachi S., Kojima M.;
 RT "Deduced hrpO genes from Burkholderia glumae and Acidovorax avenae subsp. avenae.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB053455; BAB20913.1; -;
 SQ SEQUENCE 698 AA; 75549 MW; 4B8BE407FCE559A3 CRC64;

Query Match 20.7%; Score 55.5; DB 2; Length 698;
 Best Local Similarity 27.8%; Pred. No. 48;
 Matches 15; Conservative 10; Mismatches 20; Indels 9; Gaps 1;

QY 1 MFLHSSPFKYPHTQEAQKQSRISGEMPG 45
 ID ID Q9E138 PRELIMINARY; PRT; 168 AA.
 AC Q9E138;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE POLYPROTEIN (FRAGMENT).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Pestivirus.
 OX NCBI_TaxID=121864;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-26-V639;
 RA Vilcek S., Paton D.J., Durkovic B., Strojny L., Ibata G., Moussa A.,
 RA Loitsch A., Rossmannith W., Vega S., Scicluna M.T., Palfi V.;
 RT "Bovine viral diarrhoea virus genotype 1 can be separated into at least eleven genetic groups.";
 RL Arch. Virol. 0:0-0(2000).
 DR EMBL: AF287281; AAG24564.1; -;
 FT NON_TER 168
 SQ SEQUENCE 168 AA; 18947 MW; F69B9DED96BA2C66 CRC64;

RESULT 13
 Q9E138

ID ID Q9E138 PRELIMINARY; PRT; 168 AA.
 AC Q9E138;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE POLYPROTEIN (FRAGMENT).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Pestivirus.
 OX NCBI_TaxID=121864;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-26-V639;
 RA Vilcek S., Paton D.J., Durkovic B., Strojny L., Ibata G., Moussa A.,
 RA Loitsch A., Rossmannith W., Vega S., Scicluna M.T., Palfi V.;
 RT "Bovine viral diarrhoea virus genotype 1 can be separated into at least eleven genetic groups.";
 RL Arch. Virol. 0:0-0(2000).
 DR EMBL: AF287281; AAG24564.1; -;
 FT NON_TER 168
 SQ SEQUENCE 168 AA; 18947 MW; F69B9DED96BA2C66 CRC64;

Query Match 20.5%; Score 55; DB 14; Length 168;
 Best Local Similarity 40.5%; Pred. No. 12;
 Matches 15; Conservative 4; Mismatches 14; Indels 4; Gaps 2;

QY 3 LHSSPFKYPHTQEAQKQSRISGEMPG 36
 ID ID Q9VKF5 PRELIMINARY; PRT; 196 AA.
 AC Q9VKF5;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)

RESULT 14
 Q9VKF5

ID ID Q9VKF5 PRELIMINARY; PRT; 196 AA.
 AC Q9VKF5;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)

GN	CRK.	
OS	Leishmania mexicana.	
OC	Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania	
OX	NCB_I_TaxID=5665;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=MNYC/BZ/62/M379;	
RA	Wiese M.;	
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=MNYC/BZ/62/M379;	
RA	Martin W., Wang Q., Goercke I.;	
RT	"Identification of mitogen-activated protein (MAP) kinases from	
RT	Leishmania mexicana.";	
RT	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AJ293288; CAC07964.1; -.	
KW	Kinase.	
SQ	SEQUENCE 469 AA; 52184 MW; C17DFD515BE2FD85 CRC64;	
Query Match 20.5%; Score 55; DB 5; Length 469;		
Best Local Similarity 29.8%; Pred. No. 36;		
Matches 14; Conservative 7; Mismatches 18; Indels 8; Gaps:		
Qy	1 MFLLHISPK-----YPHTQEAKQKAQRSLGEMPGRHLGSSMSL 39	
	: : : : : : :	
Db	287 MFLLGSPTRETWPSMPSYPGTWMERLASRAAAERAPDLRFGSDVTV 333	

Search completed: July 30, 2001, 16:53:06
Job time: 186 sec

Query Match 20.5%; Score 55; DB 5; Length 196;
Best Local Similarity 61.1%; Pred. No. 14;
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy	24	SLGEMPGRHLGSSMLAL	41
		: : : :	
Db	167	SIGEMPDRLHISTSSVAV	184
RESULT	15		
QGRTS			
ID	Q9GRT5	PRELIMINARY;	
AC	Q9GRT5;	PRT;	469 AA.
DT	01-WAR-2001 (Tremblrel. 16,	Created)	
DT	01-WAR-2001 (Tremblrel. 16,	Last sequence update)	
DT	01-WAR-2001 (Tremblrel. 16,	Last annotation update)	
DE	CYCLIN-DEPENDENT KINASE.		

RESULT 15
Q9GRT5
ID Q9GRT5
PRELIMINARY;
PRT; 469 AA.

DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE CYCLIN-DEPENDENT KINASE:

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model
Run on: July 30, 2001, 16:50:20 ; Search time 17.83 Seconds
(without alignments)
97.983 Million cell updates/sec

Title: US-09-402-713a-2.
Perfect score: 268
Sequence: 1 MFLHISSPKYPHTQEAQK.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.5	22.6	343	1 TA2R_BOVIN	Q95125 bos taurus
2	59	22.0	1113	1 HDA5_MOUSE	Q922v6 mus musculus
3	59	22.0	1122	1 HDA5_HUMAN	Q9uql6 homo sapien
4	55	20.5	1039	1 ITA4_MOUSE	Q00651 mus musculus
5	54.5	20.3	461	1 CO1A_BOVIN	Q92176 bos taurus
6	54.5	20.3	461	1 CO1A_HUMAN	P31146 homo sapien
7	54.5	20.3	461	1 CO1A_MOUSE	O89053 mus musculus
8	53.5	20.0	446	1 GNTU_ECOLI	P46858 escherichia
9	53.5	20.0	451	1 PTCC_BACST	Q45400 bacillus st
10	53.5	20.0	580	1 5WTD_BOOMI	P52307 boophilus m
11	53	19.8	554	1 DCS4_GOSAR	O49853 gossypium a
12	52.5	19.6	441	1 RXRA_BRARE	Q90416 brachydanio
13	52	19.4	337	1 RIR2_TRYBB	O15910 trypanosoma
14	51.5	19.2	2774	1 MAPA_RAT	P34926 rattus norv
15	51.5	19.2	269	1 AQPI_HUHO	P27568 prochloroth
16	51.5	19.2	470	1 RBL_PROHO	P29972 homo sapien
17	51.5	19.2	569	1 4CL2_ORYSA	Q42982 oryza sativ
18	51	19.0	352	1 KE4_BRARE	Q9pub8 brachydanio
19	51	19.0	1038	1 ITA4_HUMAN	P13612 homo sapien
20	51	19.0	1162	1 ITAD_HUMAN	Q13349 homo sapien
21	51	19.0	2485	1 POLN_EEVV3	P36327 venezuelan
22	51	19.0	2492	1 POLN_EEVVP	P36328 venezuelan
23	50.5	18.8	409	1 AXIA_BRARE	Q07342 brachydanio
24	50.5	18.8	554	1 DCS2_GOSAR	Q39760 gossypium a
25	50.5	18.8	2472	1 NCR2_MOUSE	Q9wu42 mus musculus
26	50	18.7	130	1 ITP_SCHGR	P26491 schistocerc
27	50	18.7	134	1 ITP_L_SCHGR	Q26490 schistocerc
28	50	18.7	456	1 HMFT_DROHY	P48590 drosophila
29	50	18.7	467	1 RBL_PHORE	P28262 phoenix rec
30	50	18.7	528	1 ANW3_RAT	O70467 rattus norv
31	50	18.7	1100	1 JAK3_RAT	P63272 rattus norv
32	50	18.7	1163	1 ITAX_HUMAN	P20702 homo sapien
33	50	18.7	2492	1 POLN_EEVVT	P27282 venezuelan

RESULT 1					
TA2R_BOVIN					
ID	TA2R_BOVIN	STANDARD;	PRT;	343 AA.	
AC	Q95125; Q95124;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	THROMBOXANE A2 RECEPTOR (TXA2-R) (PROSTANOID TP RECEPTOR).				
GN	TXA2R.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_TaxID=9913;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Liver, and Heart;				
RX	MEDLINE=98221872; PubMed=9561101;				
RA	Muck S., Schroer K.;				
RT	"Cloning, tissue-specific expression and regulation of the bovine				
RT	thromboxane A2 receptor."				
RL	Adv. Exp. Med. Biol. 433:47-50(1997).				
CC	-1- FUNCTION: RECEPTOR FOR THROMBOXANE A2 (TXA2), A POTENT STIMULATOR				
CC	OF PLATELET AGGREGATION. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED				
CC	BY A G-PROTEIN THAT ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND				
CC	MESSANGER SYSTEM. IN THE KIDNEY, THE BINDING OF TXA2 TO GLOMERULAR				
CC	TP RECEPTORS CAUSES INTENSE VASOCONSTRICTION.				
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.				
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
EMBL; U53485; AAC34309.1; -					
EMBL; U53484; AAC34308.1; -					
GCRDb; GCR1217; -					
DR	InterPro; IPR000276; -				
DR	InterPro; IPR001105; -				
DR	Pfam; PF00001; 7tm_1; 1.				
DR	PRINTS; PR00429; THROMBOXANER.				
DR	PROSITE; PS00237; G-PROTEIN RECF_F1_1; 1.				
DR	PROSITE; PS0262; G-PROTEIN RECF_F1_2; 1.				
KW	G-protein coupled receptor; Transmembrane; Glycoprotein.				
FT	DOMAIN 1 29				
FT	TRANSMEM 30 52				
FT	DOMAIN 53 66				
FT	TRANSMEM 67 87				
FT	TRANSMEM 88 106				
FT	DOMAIN 107 128				
FT	TRANSMEM 129 149				
FT	DOMAIN 150 172				
FT	TRANSMEM 150 172				

34	49.5	18.5	490	1	CPCC_RAT	P11510 rattus norv
35	49.5	18.5	499	1	RBP2_METVA	P41558 methanococc
36	49.5	18.5	554	1	DCS1_GOSHI	P93665 gossypium h
37	49.5	18.5	741	1	IDH2_VIBAL	P41561 vibrio sp.
38	49.5	18.5	893	1	PER_PERAM	Q25637 periplaneta
39	49	18.3	221	1	SCAB_SHEEP	O62816 ovis aries
40	49	18.3	405	1	PE21_RAT	P70597 rattus norv
41	49	18.3	608	1	GLMS_ECOLI	P17169 e glucosami
42	48.5	18.1	548	1	ILVG_ECOLI	P00892 escherichia
43	48.5	18.1	554	1	DCS1_GOSAR	Q39761 gossypium a
44	48.5	18.1	644	1	RNR_HELPJ	Q92jx9 helicobacte
45	48.5	18.1	834	1	CASL_HUMAN	Q14511 homo sapien

ALIGNMENTS


```
KW Hydrolase; Nuclear protein.
FT DOMAIN 47 52 POLY-GLY.
FT DOMAIN 85 92 POLY-GLN.
FT DOMAIN 596 599 POLY-GLU.
FT DOMAIN 1099 1104 POLY-ALA.
FT DOMAIN 684 1028 HISTONE DEACETYLASE.
FT CONFLICT 671 671 S -> N (IN REF. 2).
FT CONFLICT 684 684 G -> S (IN REF. 2).
FT CONFLICT 1026 1026 E -> K (IN REF. 2).
FT CONFLICT 1074 1074 E -> G (IN REF. 2).
FT CONFLICT 1085 1085 T -> M (IN REF. 2).
SQ SEQUENCE 1122 AA; 121991 MW; CF4BBB8B9A288FEC CRC64;

Query Match 22.0%; Score 59; DB 1; Length 1122;
Best Local Similarity 32.0%; Pred. No. 8.7;
Matches 16; Conservative 14; Mismatches 16; Indels 4; Gaps 2;

QY 4 HISSPFKYPHTQAEQAQRSL---GEMPRHLGSSMSIALCLVPLVREG 50
DB 386 HLTASPKLSTQAEARQALQSLRGGLTGKFMST-SIPGCLLGVALEG 434

RESULT 4
IT44_MOUSE
ID IT44_MOUSE STANDARD; PRT; 1039 AA.
IC Q00651;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE INTEGRIN ALPHA-4 PRECURSOR (INTEGRIN ALPHA-IV) (VLA-4) (CD49D)
DE (LYMPHOCYTE-Peyer's PATCH ADHESION MOLECULES ALPHA SUBUNIT) (LPAM
DE ALPHA SUBUNIT).
DE ITGA4.
GN ITGA4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92064645; PubMed=1840602;
RA Neuhaus H., Hu M.C.-T., Hemler M.E., Takada Y., Holzmann B.,
RA Weissman I.L.;
RT Cloning and expression of cDNAs for the alpha subunit of the murine
RT lymphocyte-Peyer's patch adhesion molecule.";
RL J. Cell Biol. 115:1149-1158(1991).
CC -1- FUNCTION: INTEGRINS ALPHA-4/BETA-1 (VLA-4 OR LPAM-2) AND ALPHA-
CC 4/BETA-7 (LPAM-1) ARE RECEPTORS FOR FIBRONECTIN. THEY RECOGNIZE
CC ONE OR MORE DOMAINS WITHIN THE ALTERNATIVELY SPLICED CS-1 AND CS-5
CC REGIONS OF FIBRONECTIN. THEY ARE ALSO RECEPTORS FOR VCAM1.
CC INTEGRIN ALPHA-4/BETA-1 RECOGNIZES THE SEQUENCE Q-I-D-S IN VCAM1.
CC INTEGRIN ALPHA-4/BETA-7 IS ALSO A RECEPTOR FOR MADCAM1. IT
CC RECOGNIZES THE SEQUENCE L-D-T IN MADCAM1. ON ACTIVATED ENDOTHELIAL
CC CELLS INTEGRIN VLA-4 TRIGGERS HOMOTYPIC AGGREGATION FOR MOST VLA-
CC 4-POSITIVE LEUKOCYTE CELL LINES. IT MAY ALSO PARTICIPATE IN
CC CYTOLYTIC T-CELL INTERACTIONS WITH TARGET CELLS. INTEGRIN ALPHA-
CC 4/BETA-7 IS ALSO A RECEPTOR FOR MADCAM1. MICE HOMOZYGOUS FOR A
CC NULL MUTATION OF THE ALPHA-4 SUBUNIT GENE DIE BY DAY E11-E14 FROM
CC DETACHMENT AND RUPTURE OF THE EPICARDIUM AND CORONARY ARTERIES.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC SUBUNIT CAN SOMETIMES BE CLEAVED INTO TWO NON-COVALENTLY
CC ASSOCIATED FRAGMENTS. ALPHA-4 ASSOCIATES WITH EITHER BETA-1 OR
CC BETA-7.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: PEYER'S PATCH HOMING CELLS.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X53176; CAA37316.1; -.
DR PIR; A41131; A41131.
DR PIR; S16742; S16742.
DR HSSP; P11215; IABX.
DR MGD; MGI:96603; Itga4.
DR InterPro; IPR000413; -.
DR Pfam; PF01839; FG-GAP 5.
DR Pfam; PF00357; Integrin_A; 1.
DR PRINTS; PR01185; INTEGRIN_ALPHA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
Signal; Extracellular matrix; Cytoskeleton; Repeat; Calcium.
FT SIGNAL 1 40 POTENTIAL.
FT CHAIN 41 1039 INTEGRIN ALPHA-4.
FT DOMAIN 41 983 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 984 1007 POTENTIAL.
FT DOMAIN 1008 1039 CYTOPLASMIC (POTENTIAL).
FT REPEAT 56 118 FG-GAP 1.
FT REPEAT ? ? FG-GAP 2.
FT REPEAT 193 254 FG-GAP 3.
FT REPEAT 255 308 FG-GAP 4.
FT REPEAT 310 371 FG-GAP 5.
FT REPEAT 372 431 FG-GAP 6.
FT REPEAT 434 486 FG-GAP 7.
FT CA_BIND 321 329 POTENTIAL.
FT CA_BIND 384 396 POTENTIAL.
FT CA_BIND 446 454 POTENTIAL.
FT SITE 598 599 CLEAVAGE.
FT SITE 1010 1014 GFEKR MOTIF.
FT DISULFID 98 108 BY SIMILARITY.
FT DISULFID 151 172 BY SIMILARITY.
FT DISULFID 190 205 BY SIMILARITY.
FT DISULFID 493 502 BY SIMILARITY.
FT DISULFID 508 564 BY SIMILARITY.
FT DISULFID 629 634 BY SIMILARITY.
FT DISULFID 705 718 BY SIMILARITY.
FT DISULFID 859 897 BY SIMILARITY.
FT DISULFID 904 909 BY SIMILARITY.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 145 145 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 236 236 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 487 487 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 525 525 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 545 545 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 652 652 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 813 813 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 828 828 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1039 AA; 115695 MW; C8C0BC611F7E9847 CRC64;

Query Match 20.5%; Score 55; DB 1; Length 1039;
Best Local Similarity 28.3%; Pred. No. 27;
Matches 17; Conservative 6; Mismatches 19; Indels 18; Gaps 2;

QY 9 FKYPHTQAEQAQRSL---QRSIG-----EMPRHLGSSMSIALCLVPLVREG 50
DB 267 FRSPHTTEVVGGAPOEHOIGKAYIFSIDENELNIVEMKGLSGVFGASCAVDLNADG 326

RESULT 5
ID CO1A_BOVIN STANDARD; PRT; 461 AA.
AC 092176;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CORONIN-LIKE PROTEIN P57 (CORONIN 1A).
```

GN CORO1A OR CORO1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Eutheria; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Spleen;
 RX MEDLINE=95278344; PubMed=7758584;
 RA Suzuki K., Nishihata J., Arai Y., Honma N., Yamamoto K., Irimura T.,
 RA Toyoshima S.;
 RT "Molecular cloning of a novel actin-binding protein, p57, with a WD
 RT repeat and a leucine zipper motif.";
 RL FEBS Lett. 364:283-288(1995).
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF
 CC CHEMOTAXIS.
 CC -!- SUBUNIT: BINDS ACTIN.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, THYMUS, SPLEEN, BONE
 CC MARROW AND LYMPH NODE. LOW IN LUNG AND GUT.
 CC -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
 CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC EMBL; D44496; BAA07939.1;
 CC InterPro: IPR001680;
 CC Pfam: PF00400; WD40; 3.
 CC PROSITE; PS00678; WD_REPEATS_1; 2.
 CC PROSITE; PS50082; WD_REPEATS_2; 2.
 CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
 CC Actin-binding; Repeat; WD repeat; Coiled coil.
 FT REPEAT 79 119 WD 1.
 FT REPEAT 129 169 WD 2.
 FT REPEAT 174 213 WD 3.
 FT REPEAT 218 260 WD 4.
 FT REPEAT 265 305 WD 5.
 FT DOMAIN 424 460 COILED COIL (POTENTIAL).
 FT SEQUENCE 461 AA; 50979 MW; 1721A5D093C1130A CRC64;
 SQ
 Query Match 20.3%; Score 54.5; DB 1; Length 461;
 Best Local Similarity 43.8%; Pred. No. 13;
 Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;
 OY 2 FLHISPPKYPHTQEAQKQASLGEMPGRHL 33
 DB 303 FLHLSMF-----SSKESQRMGYMPKRG 327
 RESULT 6
 CO1A.HUMAN
 ID CO1A.HUMAN STANDARD; PRT; 461 AA.
 AC P31146;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE CORONIN-LIKE PROTEIN P57 (CORONIN 1A).
 GN CORO1A OR CORO1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Peripheral blood leukocytes;
 RX MEDLINE=95278344; PubMed=7758584;

RA Suzuki K., Nishihata J., Arai Y., Honma N., Yamamoto K., Irimura T.,
 RA Toyoshima S.;
 RT "Molecular cloning of a novel actin-binding protein, p57, with a WD
 RT repeat and a leucine zipper motif.";
 RL FEBS Lett. 364:283-288(1995).
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF
 CC CHEMOTAXIS.
 CC -!- SUBUNIT: BINDS ACTIN (BY SIMILARITY).
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, THYMUS, SPLEEN, BONE
 CC MARROW AND LYMPH NODE. LOW IN LUNG AND GUT.
 CC -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
 CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC EMBL; D44497; BAA07940.1;
 CC EMBL; X89109; CAA61482.1;
 CC EMBL; U34690; AAA77058.1;
 CC MIM; 605000;
 CC Aarhus/Ghent-2DPAGE; 2416; IEF.
 CC InterPro: IPR001680;
 CC Pfam: PF00400; WD40; 3.
 CC PROSITE; PS00678; WD_REPEATS_1; 2.
 CC PROSITE; PS50082; WD_REPEATS_2; 2.
 CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
 CC Actin-binding; Repeat; WD repeat; Coiled coil.
 FT REPEAT 79 119 WD 1.
 FT REPEAT 129 169 WD 2.
 FT REPEAT 174 213 WD 3.
 FT REPEAT 218 260 WD 4.
 FT REPEAT 265 305 WD 5.
 FT DOMAIN 424 460 COILED COIL (POTENTIAL).
 FT VARIANT 355 355 K->D.
 FT /FTID=VAR_007722.
 FT S->D.
 FT /FTID=VAR_007723.
 FT S->T (IN REF. 3).
 FT CONFLICT 245 245 R->W (IN REF. 3).
 FT SEQUENCE 461 AA; 51026 MW; DE3FEDA57041515E CRC64;
 SQ
 Query Match 20.3%; Score 54.5; DB 1; Length 461;
 Best Local Similarity 43.8%; Pred. No. 13;
 Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;
 OY 2 FLHISPPKYPHTQEAQKQASLGEMPGRHL 33
 DB 303 FLHLSMF-----SSKESQRMGYMPKRG 327

```

RESULT 7
COIA_MOUSE STANDARD; PRT; 461 AA.
ID COIA_MOUSE STANDARD; PRT; 461 AA.
AC O89053; Q9R288;
DT 30-MAY-2000 (Rel. 39, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CORONIN-LIKE PROTEIN P57 (CORONIN 1A).
GN COROLA OR COROL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449467; PubMed=9778037;
RA Okumura M., Kung C., Wong S., Rodgers M., Thomas M.L.;
RT "Definition of family of coronin-related proteins conserved between
RT humans and mice: close genetic linkage between coronin-2 and CD45-
RT associated protein."
RL DNA Cell Biol. 17:779-787(1998).
RN [2]
RP SEQUENCE OF 156-276 FROM N.A.
RX STRAIN=BALE/C; TISSUE=Spleen;
RX MEDLINE=99012997; PubMed=9798653;
RA Chu C.C., Paul W.E.;
RT "Expressed genes in interleukin-4 treated B cells identified by cDNA
RT representational difference analysis."
RL Mol. Immunol. 35:487-502(1998).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF
CC CHEMOTAXIS (BY SIMILARITY).
CC -!- SUBUNIT: BINDS ACTIN (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSION RESTRICTED TO HEMATOPOIETIC CELLS.
CC -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF143955; AAD32703.1;
DR EMBL; U89399; AAC36506.1;
DR MGD; MGI:1345961; Corola.
DR InterPro; IPR001680;
DR Pfam; PF00400; WD40; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS00082; WD_REPEATS_2; 2.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Actin-binding; Repeat; WD repeat; Coiled coil.
FT REPEAT 79 119 WD 1.
FT REPEAT 129 169 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 218 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 424 460 COILED COIL (POTENTIAL).
FT SEQUENCE 461 AA; 50975 MW; 51D48D91E6FE82A CRC64;

Query Match 20.3%; Score 54.5; DB 1; Length 461;
Best Local Similarity 43.8%; Pred. NO. 13;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

QY 2 FLHISPPFYKPTQEAQKQSRSLGEMPCRHLL 33
DB 303 FLHYSMF-----SSKESQRMGMYPKRGRL 327

```

```

RESULT 8
GNTU_ECOLI STANDARD; PRT; 445 AA.
ID GNTU_ECOLI STANDARD; PRT; 445 AA.
AC P46858; P76694; P76695;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE LOW-AFFINITY GLUCONATE TRANSPORTER (GLUCONATE PERMEASE) (GNT-I
DE SYSTEM).
GN GNTU OR B3435/B3436.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / W3110;
RX MEDLINE=97280784; PubMed=9135111;
RA Izu H., Adachi O., Yamada M.;
RT "Gene organization and transcriptional regulation of the gntRku
RT operon involved in gluconate uptake and catabolism of Escherichia
RT coli."
RL J. Mol. Biol. 267:778-793(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=96235044; PubMed=8655507;
RA Tong S., Porco A., Istriz T., Conway T.;
RT "Cloning and molecular genetic characterization of the Escherichia
RT coli gntR, gntK, and gntU genes of GntI, the main system for
RT gluconate metabolism."
RL J. Bacteriol. 178:3260-3269(1996).
CC -!- FUNCTION: PART OF THE GLUCONATE UTILIZATION SYSTEM GNT-I; LOW-
CC AFFINITY INTAKE OF GLUCONATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE GNTP FAMILY OF PERMEASES.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D84362; BAAL2326.1;
DR EMBL; U18997; AAA58234.1; ALT_FRAME.
DR EMBL; U18997; AAA58233.1; ALT_FRAME.
DR EMBL; AE000420; AAC76461.1; ALT_FRAME.
DR EMBL; AE000420; AAC76460.1; ALT_FRAME.
DR EcoGene; EG12631; gntU.
KW Gluconate utilization; Sugar transport; Transmembrane;
KW Inner membrane; Multigene family.
FT TRANSMEM 2 22 POTENTIAL.
FT TRANSMEM 27 47 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 110 130 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.

```



```
FT CONFLICT 15 15 T -> K (IN REF. 2).
FT CONFLICT 37 39 SGT -> HXG (IN REF. 2).
SQ SEQUENCE 580 AA; 63460 MW; 588EEF2014071AB7 CRC64;

Query Match
Best Local Similarity 36.1%; Score 53.5; DB 1; Length 580;
Matches 22; Conservative 5; Mismatches 21; Indels 13; Gaps 3;

Qy 4 HISSPFKYPHTQEA--OKEAQRSLGE-----MPGRHLGSSMSLALC-----LYPLVREG 50
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 175 YLSSPGKVRFTDEACIQEAQRRLREGVQVIAVGHSGVPRDLCEICRERPEVSLVVG 234
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 51 H 51
   |
Db 235 H 235

RESULT 11
DCS4_GOSAR STANDARD; PRT; 554 AA.
AC 049853; 1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE (+)-DELTA-CADINENE SYNTHASE ISOZYME C2 (EC 4.6.1.11) (D-CADINENE
SYNTHASE).
GN CAD1-C2.
OS Gossypium arboreum.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
OX NCBI_TaxID=29729;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NANKING;
RA Meng Y., Jia J., Liu C., Liang W., Zhou X., Heinstein P., Chen X.-Y.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: (E,E)-FARNESYL DIPOSPHATE CYCLIZING.
CC -1- CATALYTIC ACTIVITY: 2-TRANS-6-TRANS-FARNESYL DIPOSPHATE = (+)-
DELTA-CADINENE + DIPHOSPHATE.
CC -1- PATHWAY: FIRST COMMITTED STEP IN THE PATHWAYS LEADING TO THE
RELATED PHYTOALEXINS GOSSYPOL AND LACINILENE C.
CC -1- SIMILARITY: TO OTHER PLANT TERPENE CYCLASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; Y16432; CAA76223.1; -.
CC HSSP; Q40577; 5EAU.
CC InterPro; IPR001906; -.
CC Pfam; PF01397; Terpene_synth; 1.
CC Lyase; Multigene family.
SQ SEQUENCE 554 AA; 64117 MW; 35DD666D3E838AAC CRC64;

Query Match
Best Local Similarity 38.3%; Score 53; DB 1; Length 554;
Matches 18; Conservative 5; Mismatches 18; Indels 6; Gaps 2;

Qy 4 HISSPFKYPHTQEAQRSLGEMPGRHLGSSMSLALCLVPLVREG 50
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 490 HVESAWK-DVNEQFKPT-----EMPTVLRSLNLRVMDVLYREG 530
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 12
RXRA_BRARE STANDARD; PRT; 441 AA.
ID RXRA_BRARE
```

```
AC Q90416;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RETINOIC ACID RECEPTOR RXR-ALPHA.
GN RXRA OR NR2B1 OR RXR.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96009547; PubMed=7565671;
RA Jones B.B., Ohno C.K., Allenby G., Boffa M.B., Levin A.A.,
RA Grippo J.F., Petkovich M.;
RT "New retinoid X receptor subtypes in zebra fish (Danio rerio)
RT differentially modulate transcription and do not bind 9-cis retinoic
RT acid.";
RL Mol. Cell. Biol. 15:5226-5234(1995).
CC -1- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS
9-CIS RETINOIC ACID (9C-RA) (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
NR2 SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; U29940; AAC59720.1; -.
CC HSSP; P19793; 2NLL.
CC ZFIN; ZDB-GENE-980526-36; rxra.
CC InterPro; IPR000003; -.
CC InterPro; IPR000536; -.
CC InterPro; IPR001628; -.
CC InterPro; IPR001723; -.
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00047; STROIDFINGER.
CC PRINTS; PR00398; STRDHORMONER.
CC PRINTS; PR00545; RETINOIDXR.
CC PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family.
FT DOMAIN 1 116 MODULATING (BY SIMILARITY).
FT DNA_BIND 117 182 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 117 137 C4-TYPE.
FT ZN_FING 153 177 C4-TYPE.
FT DOMAIN 183 206 HINGE.
FT DOMAIN 207 441 LIGAND-BINDING (BY SIMILARITY).
SQ SEQUENCE 441 AA; 48690 MW; E0F8F1EBE31CA1EC CRC64;

Query Match
Best Local Similarity 31.1%; Score 52.5; DB 1; Length 441;
Matches 14; Conservative 6; Mismatches 18; Indels 7; Gaps 1;

Qy 2 FLHISSPFKYPHTQEAQRSLGEMPGRH-----LGSSMSL 39
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 7 YLHLSSSLQVAHGLSLSPSPQPLGLSMVSHHHPHSINGLGLSPSY 51
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

RESULT 13
RXR2_TRYBB STANDARD; PRT; 337 AA.
ID RXR2_TRYBB
```


AC 015910; 015880;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (EC 1.17.4.1)
DE (RIBONUCLEOTIDE REDUCTASE R2 SUBUNIT).
GN RN2 OR NRDB.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN (1)
RN SEQUENCE FROM N.A.
RP STRAIN=427;
RX MEDLINE=97338127; PubMed=9192674;
RA Hofer A., Schmidt P.P., Graslund A., Thelander L.;
RT "Cloning and characterization of the R1 and R2 subunits of
RT ribonucleotide reductase from Trypanosoma brucei";
RL Proc. Natl. Acad. Sci. U.S.A. 94:6959-6964(1997).
RN [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=97459778; PubMed=9315738;
RX Dornmeyer M., Schoneck R., Dittmar G.A.G., Krauth-Siegel R.L.;
RA "Cloning, sequencing and expression of ribonucleotide reductase R2
RT from Trypanosoma brucei";
RT FEBS Lett. 414:449-453(1997).
CC -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC -!- CATALYTIC ACTIVITY: 2'DEoxyRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC THIOREDOXIN + H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC THIOREDOXIN.
CC -!- COFACTOR: CONTAINS TWO IRON IONS.
CC -!- PATHWAY: FIRST REDIMER IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC SMALL CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U80911; AAB70705.1; -
DR EMBL; Y10768; CAA71741.1; -
DR InterPro; IPR000358; -
DR Pfam; PF00268; ribonuc_red.1.
DR PROSITE; PS00368; RIBRED_SMALL; 1.
KW Oxidoreductase; DNA replication; Iron.
FT METAL 85 85
FT METAL 116 116 IRON 1 (BY SIMILARITY).
FT METAL 119 119 IRON 1 AND 2 (BY SIMILARITY).
FT METAL 178 178 IRON 1 (BY SIMILARITY).
FT METAL 178 178 IRON 2 (BY SIMILARITY).
FT METAL 212 212 IRON 2 (BY SIMILARITY).
FT METAL 215 215 IRON 2 (BY SIMILARITY).
FT ACT_SITE 123 123 BY SIMILARITY.
FT CONFLICT 8 10 CSR -> RSA (IN REF. 2).
FT CONFLICT 43 43 E -> K (IN REF. 2).
SQ SEQUENCE 337 AA; 39008 MW; BD39919280C65CAE CRC64;

Query Match 19.4%; Score 52; DB 1; Length 337;
Best Local Similarity 31.4%; Pred. No. 20;
Matches 11; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 4 HISSPKYPTQEAQKQSRSLGEMPGRHLSMS 38
DB :: |||| : ||| : : ||| :
27 YVIFPKYPTQWQYKEASSIWTVEIDLGNMT 61

RESULT 14
ID MAPA_RAT STANDARD; PRT: 2774 AA.

AC P34926;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2].
GN MAP1A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=92355629; PubMed=1379599;
RA Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
RT one messenger RNA";
RL J. Biol. Chem. 267:16561-16566(1992).
CC -!- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN THE FILAMENTOUS
CC CROSS-BRIDGING BETWEEN MICROTUBULES AND OTHER SKELETAL ELEMENTS.
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
CC APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
CC THEIR MORPHOLOGY.
CC -!- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
CC FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC -!- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B.
CC -!- SIMILARITY: TO MAP1B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M83196; AAB48069.1; -
DR PIR; A43359; A43359.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN 72465 2774 MAP1 LIGHT CHAIN LC2.
FT DOMAIN 309 496 LYS-RICH (BASIC).
FT DOMAIN 336 541 11 X 3 AA REPEATS OF K-K-[DE].
FT REPEAT 336 338 1.
FT REPEAT 415 417 2.
FT REPEAT 420 422 3.
FT REPEAT 424 426 4.
FT REPEAT 427 429 5.
FT REPEAT 431 433 6.
FT REPEAT 436 438 7.
FT REPEAT 440 442 8.
FT REPEAT 444 446 9.
FT REPEAT 449 451 10.
FT REPEAT 539 541 11.
SQ SEQUENCE 2774 AA; 299526 MW; 3DEF74427BA9D7D7 CRC64;

Query Match 19.4%; Score 52; DB 1; Length 2774;
Best Local Similarity 46.2%; Pred. No. 1.8e+02;
Matches 12; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 9 FKYPHTQEAQKQSRSLGEMPGRHLS 34
DB ||||| : : ||| :
678 FYQKHTQEAQKQSRSLGEMPGRHLS 703

RESULT 15
AQPI_HUMAN

AD AQP1_HUMAN STANDARD; PRT; 269 AA.
AC P29972;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AQPAPORIN-CHIP (WATER CHANNEL PROTEIN FOR RED BLOOD CELLS AND KIDNEY
DE PROXIMAL TUBULE) (AQPAPORIN 1) (AQP-1) (URINE WATER CHANNEL).
GN AQP1 OR CHIP28.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92107900; PubMed=1722319;
RA Preston G.M., Agre P.;
RT "Isolation of the cDNA for erythrocyte integral membrane protein of
RT 28 kilodaltons: member of an ancient channel family.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:11110-11114(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93340184; PubMed=8340403;
RA Moon C., Preston G.M., Griffin C.A., Jabs E.W., Agre P.;
RT "The human aquaporin-CHIP gene. Structure, organization, and
RT chromosomal localization.";
RL J. Biol. Chem. 268:15772-15778(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Retinal pigment epithelium;
RC MEDLINE=96326579; PubMed=8703970;
RA Ruiz A.C., Bok D.;
RT "Characterization of the 3' UTR sequence encoded by the AQP-1 gene in
RT human retinal pigment epithelium.";
RL Biochim. Biophys. Acta 1282:174-178(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=94290349; PubMed=7517253;
RA Li X., Yu H., Koide S.S.;
RT "The water channel gene in human uterus.";
RL Biochem. Mol. Biol. Int. 32:371-377(1994).
RN [5]
RP SEQUENCE OF 1-128 FROM N.A.
RA Dempsey S., Lacy M., Holmes A., Nguyen C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 129-269 FROM N.A.
RA Andrews S., Dubbelde C., Ryan E.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP FUNCTION.
RX MEDLINE=92229472; PubMed=1373524;
RA Preston G.M., Carroll T.P., Guggino W.B., Agre P.;
RT "Appearance of water channels in Xenopus oocytes expressing red cell
RT CHIP28 protein.";
RL Science 256:385-387(1992).
RN [8]
RP TARGET OF MERCURY INHIBITION.
RX MEDLINE=93106996; PubMed=7677994;
RA Preston G.M., Jung J.S., Guggino W.B., Agre P.;
RT "The mercury-sensitive residue at cysteine 189 in the CHIP28 water
RT channel.";
RL J. Biol. Chem. 268:17-20(1993).
RN [9]
RP TOPOLOGY.
RX MEDLINE=94124503; PubMed=7507481;
RA Preston G.M., Jung J.S., Guggino W.B., Agre P.;
RT "Membrane topology of aquaporin CHIP. Analysis of functional epitope-
RT scanning mutants by vectorial proteolysis.";
RL J. Biol. Chem. 269:1668-1673(1994).
RN [10]
RP STRUCTURE BY ELECTRON CRYO-MICROSCOPY.
RX MEDLINE=94313979; PubMed=7518771;
RA Walz T., Smith B.L., Agre P., Engel A.;
RT "The three-dimensional structure of human erythrocyte aquaporin
RT CHIP.";
RL EMBO J. 13:2985-2993(1994).
RN [11]
RP STRUCTURE BY ELECTRON CRYO-MICROSCOPY.
RX MEDLINE=97320502; PubMed=9177353;
RA Walz T., Hirai T., Murata K., Heymann J.B., Mitsuoka K.,
RA Fujiyoshi Y., Smith B.L., Agre P., Engel A.;
RT "The three-dimensional structure of aquaporin-1.";
RL Nature 387:624-627(1997).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (3.8 ANGSTROMS).
RX MEDLINE=20487015; PubMed=11034202;
RA Murata K., Mitsuoka K., Hirai T., Walz T., Agre P., Heymann J.B.,
RA Engel A., Fujiyoshi Y.;
RT "Structural determinants of water permeation through aquaporin-1.";
RL Nature 407:599-605(2000).
RN [13]
RP VARIANT BLOOD GROUP COLTON.
RX MEDLINE=94365170; PubMed=7521882;
RA Smith B.L., Preston G.M., Spring F., Anstee D.J., Agre P.;
RT "Human red cell aquaporin CHIP. I. Molecular characterization of ABH
RT and Colton blood group antigens.";
RL J. Clin. Invest. 94:1043-1049(1994).
RN [14]
RP FUNCTION: FORMS A WATER-SPECIFIC CHANNEL THAT PROVIDES THE PLASMA
RP MEMBRANES OF RED CELLS AND KIDNEY PROXIMAL TUBULES WITH HIGH
RP PERMEABILITY TO WATER, THEREBY PERMITTING WATER TO MOVE IN THE
RP DIRECTION OF AN OSMOTIC GRADIENT.
RN [15]
RP SUBUNIT: HOMOTETRAMER.
RN [16]
RP SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
RN [17]
RP TISSUE SPECIFICITY: EXPRESSED IN A NUMBER OF TISSUES INCLUDING
RP ERYTHROCYTES, RENAL TUBULES, RETINAL PIGMENT EPITHELIUM,
RP HEART, LUNG, SKELETAL MUSCLE, KIDNEY AND PANCREAS. WEAKLY
RP EXPRESSED IN BRAIN, PLACENTA AND LIVER.
RN [18]
RP DOMAIN: HB AND HE TOGETHER FORM A TRANSMEMBRANE HELIX. THEY ALSO
RP FUNCTION AS PORE HELICES INVOLVED IN DEFINING SUBSTRATE
RP SPECIFICITY. THE TWO NPA REPEATS ARE SITUATED AT THE END OF HB AND
RP HE AND ARE IMPORTANT FOR THE INTERACTION BETWEEN THESE TWO
RP HELICES.
RN [19]
RP POLYMORPHISM: AQP1 IS RESPONSIBLE FOR THE COLTON BLOOD GROUP
RP SYSTEM. APPROXIMATELY 92% OF CAUCASIANS ARE CO(A+B-) (ALA-45),
RP APPROXIMATELY 8% ARE CO(A+B+), AND ONLY 0.2% ARE CO(A-B+) (VAL-
RP 45). CO(A-B-) WHICH IS VERY RARE, IS DUE TO A COMPLETE ABSENCE OF
RP AQP1.
RN [20]
RP MISCELLANEOUS: PHARMACOLOGICALLY INHIBITED BY SUBMILLIMOLAR
RP CONCENTRATIONS OF Hg2+.
RN [21]
RP SIMILARITY: BELONGS TO THE TRANSMEMBRANE CHANNEL MIP FAMILY.
RN [22]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL Outstation -
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RP or send an email to license@isb-sib.ch).
RN [23]
RP EMBL; M7829; AAA58425.1; -
DR EMBL; U41517; AAC50848.1; -
DR EMBL; U41518; AAC50849.1; -
DR EMBL; S73482; AAB31193.1; -
DR EMBL; AC004691; AAC16481.1; -
DR EMBL; AC005155; AAC23788.1; -
DR PIR; A41616; A41616.
DR PDB; 1FQY; 18-OCT-00.
DR MIM; 110450; -
DR InterPro; IPR000425; -
DR Pfam; PF00230; MIP; 1.
DR PRINTS; PR00783; MINTRINSCP.
DR PROSITE; PS00221; MIP; 1.
KW Transport; Transmembrane; Polymorphism; Blood group antigen;
KW Glycoprotein; 3D-structure.

```

FT DOMAIN 1 7 CYTOPLASMIC.
FT TRANSMEM 8 36 HELIX 1.
FT DOMAIN 37 48 EXTRACELLULAR.
FT TRANSMEM 49 66 HELIX 2.
FT DOMAIN 67 70 CYTOPLASMIC.
FT DOMAIN 71 76 IN MEMBRANE.
FT TRANSMEM 77 84 HELIX B.
FT DOMAIN 85 94 CYTOPLASMIC.
FT TRANSMEM 95 115 HELIX 3.
FT DOMAIN 116 136 EXTRACELLULAR.
FT TRANSMEM 137 155 HELIX 4.
FT DOMAIN 156 186 CYTOPLASMIC.
FT TRANSMEM 167 183 HELIX 5.
FT DOMAIN 184 186 EXTRACELLULAR.
FT TRANSMEM 187 192 IN THE MEMBRANE.
FT DOMAIN 193 200 HELIX E.
FT TRANSMEM 201 207 EXTRACELLULAR.
FT TRANSMEM 208 228 HELIX 6.
FT DOMAIN 229 269 CYTOPLASMIC.
FT REPEAT 76 78
FT REPEAT 192 194
FT DOMAIN 159 162
FT SITE 189 189
FT CARBOHYD 42 42
FT CARBOHYD 205 205
FT VARIANT 45 45
FT SEQUENCE 269 AA; 28526 MW; BA204D82FB26352E CRC64;

POLY-ARG
HG(2+)-SENSITIVE RESIDUE.
N-LINKED (GLCNAC. . .).
N-LINKED (GLCNAC. . .) (POTENTIAL).
A -> V (IN CO(A-B+) ANTIGEN).
/FTIG-VAR_004400.

```

```

Query Match 19.2%; Score 51.5; DB 1; Length 269;
Best Local Similarity 26.1%; Pred. No. 18;
Matches 18; Conservative 13; Mismatches 17; Indels 21; Gaps 4;

QY 1 MELHISP--FKYP--HTQEAQKE-----AQRSLGEMPGRHLSGSMNLAL- 41
Db 25 VFISIGSALGFKYPVGNNGTAVODNVKVSARFGLSLATLAQSVGHISGAHLNPAVTLGLL 84

QY 42 --CLVPLVR 48
Db 85 LSCQISIFR 93

```

Search completed: July 30, 2001, 16:53:30
Job time: 190 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:48:05 ; Search time 26.94 Seconds
(without alignments)
144.206 Million cell updates/sec

Title: US-09-402-713a-2
Perfect score: 268
Sequence: 1 MELHTSSPKYPTQAEK.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	62.5	23.3	441	2	E86758	dihydroorotase (EC
2	60.5	22.6	1239	2	T13809	probable disintegr
3	60	22.4	187	2	B81903	hypothetical prote
4	60	22.4	218	2	C81118	hypothetical prote
5	56.5	21.1	1061	2	S75508	acriflavin resista
6	56	20.9	338	2	C75459	probable endonucle
7	55	20.5	1039	2	A41131	lymphocyte-Peyer's
8	54.5	20.3	244	2	A83236	hypothetical prote
9	54.5	20.3	305	2	T47844	hypothetical prote
10	54.5	20.3	461	2	S65665	actin-binding prot
11	54.5	20.3	461	2	S65666	actin-binding prot
12	54	20.1	204	2	T51012	hypothetical prote
13	54	20.1	257	2	G72081	conserved hypothet
14	54	20.1	257	2	F86540	C254 hypothetical
15	54	20.1	633	2	B82990	hypothetical prote
16	53.5	20.0	446	2	T14128	NADH dehydrogenase
17	53.5	20.0	348	2	JC4989	gluconate permease
18	53.5	20.0	446	2	B86010	low affinity glucos
19	53.5	20.0	449	2	D72127	hypothetical prote
20	53.5	20.0	451	2	C49898	cellobiose phospho
21	52.5	19.6	216	2	E75446	amino acid ABC tra
22	52.5	19.6	357	2	E83632	hypothetical prote
23	52.5	19.6	441	2	I50515	retinoid X recepto
24	52.5	19.6	493	2	T43465	hypothetical prote
25	52.5	19.6	563	2	T17316	hypothetical prote
26	52	19.4	210	2	H75599	probable acetyltra
27	52	19.4	493	2	F96696	protein FIN21.12
28	52	19.4	2774	2	A43359	microtubule-associ
29	51.5	19.2	173	2	D83486	cobinamide kinase

30 51.5 19.2 200 2 H84715 probable phytoeyan
31 51.5 19.2 269 2 A41616 erythrocyte integr
32 51.5 19.2 380 2 S75929 sensory transducti
33 51.5 19.2 449 2 C86496 hypothetical prote
34 51.5 19.2 449 2 A81544 hypothetical prote
35 51.5 19.2 470 1 RKMWLX ribulose-bisphosph
36 51.5 19.2 569 2 T03390 4-coumarate--CoA 1
37 51.5 19.2 137 2 B72786 hypothetical prote
38 51.5 19.0 464 2 T45612 N-hydroxycinnamoyl
39 51.5 19.0 633 2 A75069 dipeptide transpor
40 51.5 19.0 1038 2 S06046 integrin alpha-4 c
41 51.5 19.0 2137 2 T05244 hypothetical prote
42 51.5 19.0 2492 1 A44213 nonstructural poly
43 51.5 19.0 2492 1 C44213 nonstructural poly
44 50.5 18.8 115 2 G72617 hypothetical prote
45 50.5 18.8 409 1 A47527 transcription fact

ALIGNMENTS

RESULT 1

E86758
dihydroorotase (EC 3.5.2.3) [imported] - Lactococcus lactis subsp. lactis (strain IL1
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C:Accession: E86758
R:Boletini, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh
Genome Res. in press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: E86758
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-441 <STO>
A:Cross-references: GB:AF005176; NID:g12724023; PIDN:AAK05167.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: pyrC
C:Superfamily: Bacillus dihydroorotase; Bacillus dihydroorotase homology
C:Keywords: hydrolase

Query Match 23.3%; Score 62.5; DB 2; Length 441;

Best Local Similarity 42.5%; Pred. No. 1.6;

Matches 17; Conservative 4; Mismatches 16; Indels 3; Gaps 1;

Oy 12 PHTQEAQKQKRSLEMPGRHLGSSMSLALCLVPLVREGH 51

Db 325 PHTRE---EKNVSLDKAPSGMIGLETSLQLGLTNLVAKGH 361

RESULT 2

T13809

probable disintegrin (EC 3.4.24.-) - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13809

R:Roake, J.; Pan, D.; Xu, T.; Rubin, G.M.

Science 273, 1227-1231, 1996

A:Title: KUZ, a conserved metalloprotease-disintegrin with two roles during Drosophil

A:Reference number: Z17772; MUID:96355647

A:Accession: T13809

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1239 <ROO>

A:Cross-references: EMBL:U60591; NID:g1531632; PID:g1531633; PIDN:AA47275.1

C:Genetics:

A:Gene: kuz

A:Cross-references: FlyBase:FBgn0015954

A:Map position: 2

C:Keywords: hydrolase; metalloproteinase

Query Match 22.6%; Score 60.5; DB 2; Length 1239;
Best Local Similarity 47.2%; Pred. No. 9.7;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPHQTQAEAKQRSLGEM----PGRHLGSSMSLA 40
||||| : ||| : | || |||| : :|
Db 341 FKYPH-QKYTKEANFAEGAFDPSTGRRLGSSANVA 375

RESULT 3

B81903
hypothetical protein NMA1341 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C;Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: B81903
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A811775; MUID:2022556
A:Accession: B81903
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-187 <PAR>
A:Cross-references: GB:AL157959; NID:g7379742; PIDN:CAB84590.1; PID:g738001
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA1341

```

Query Match      22.4%; Score 60; DB 2; Length 187;
Best Local Similarity 31.4%; Pred. No. 1.4;
Matches 16; Conservative 11; Mismatches 22; Indels 2; Gaps 1;

QY 1 MFLHSPPFYPHPTQAEQAKQSRSLGEMPGRHGSSMSLALCLVPLVREGH 51
   || : : : : : : : : : : : : : : : : : : : : : :
58 LFLSLRGIEIVPKNETAESEPKSVHNER--ROENAGSGRKOLLIDLVRGH 106
   || : : : : : : : : : : : : : : : : : : : : : :
Db

```

RESULT 4

C81118
hypothetical protein NMB1132, NMB1170 [imported] - Neisseria meningitidis (strain MC58 S
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81118; H81113
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.;
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
Iri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: C81118
A:Molecule type: DNA
A:Residues: 1-218 <TET>
A:Cross-references: GB:AE002462; GB:AE002098; NID:g7226363; PIDN:AAF41520.1; PID:g722636
A:Experimental source: serogroup B, strain MC58
A:Accession: H81113
A:Molecule type: DNA
A:Residues: 1-218 <TET2>
A:Cross-references: GB:AE002465; GB:AE002098; NID:g7226401; PIDN:AAF41555.1; PID:g722640
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1132; NMB1170

```
Query Match          22.4%; Score 60; DB 2; Length 218;
Best Local Similarity 31.4%; Pred. No. 1.6;
Matches 16; Conservative 11; Mismatches 22; Indels 2; Gaps 1;
```

DB 89 LFSLRGEIVFPKNETAESEKSVHER--RQEGNAGSGRKQLDLVRRGH 137

RESULT 5
S75508
acriflavin resistance protein acrf - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein slr2131
C:Species: synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S75508
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocys
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S75508
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1061 <KAN>
A:CROSS-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAA18069.1; PID:g165
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: acrf
C:Superfamily: acriflavin resistance protein

Query Match 21.1%; Score 56.5; DB 2; Length 1061;
Best Local Similarity 31.0%; Pred. No. 28;
Matches 13; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

Qy 5 ISSPFKYPTGQAQRSLGEMPGRRHLGSMSLALCLVPL 46
|| | : : || : : || : : || : : || : :
Db 418 ISRFIKVKH-EDPOEAQAQGELTGAVIATSLVLMVFIPV 458

RESULT 6
C75459
probable endonuclease III - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: C75459
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: C75459
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <WHI>
A:CROSS-references: GB:AE001945; GB:AE000513; NID:g6458642; PIDN:AAF10505.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0928
A:Map position: 1

```

Query Match      20.9%; Score 56; DB 2; Length 338;
Best Local Similarity 38.1%; Pred. No. 9.3;
Matches 8; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 10 KYPHTQEAQKEAQRSLGEMPG 30
      :| | | | :| | | :| |
Db 207 EFPHTPEGHEQAQQAALPG 227

RESULT 7
A41131
lymphocyte-Peyer's patch adhesion molecule alpha 4 precursor - mouse
N;Alternate names: integrin alpha-4

```


QY 2 FLHISSPFKYPTHQEAQKEAQRSLGEMPRHL 33
 ||| | | : ||| : | | |
 Db 303 FLHLYSMF-----SSKESQRGMGYMPKRG 327

RESULT 12

T51012
hypothetical protein B7J19.110 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 08-Sep-2000
C:Accession: T51012
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, B. submitted to the Protein Sequence Database, July 2000
A:Reference number: Z25286
A:Accession: T51012
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <SCH>
A:Cross-references: EMBL:AL389899; GSPDB:GN001116; NCSP:B7J19.110
A:Experimental source: BAC clone B7J19; strain OR74A
C:Genetics:
A:Gene: NCSP:B7J19.110
A:Map position: 6
A:introns: 31/1; 66/3; 170/3
C:Superfamily: Neurospora crassa hypothetical protein B7J19.110

```
Query Match      20.1%; Score 54; DB 2; Length 204;
Best Local Similarity 31.2%; Pred.No. 9;9;
Matches 15; Conservative 12; Mismatches 11; Indels

QY 13 HTQBAQEAKQRSL--GEMPGRHL-----GSMSIALCLVPLVLREGH 51
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 153 HSKQ-KRETYITHNGQLGGHLVPQKALGGSGASNCWCVLAQHFI 199
```

RESULT 13

G72081 conserved hypothetical protein CP0355 [Imported] - Chlamydothrix pneumoniae (strains CWT
 N:Alternate names: ct254 hypothetical protein
 C:Species: Chlamydothrix pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 18-Aug-2000
 C:Accession: G72081; H81584
 R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; J.
 Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606
 A:Accession: G72081
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-257 <ARN>
 A:Cross-references: GB:AE001624; GB:AE001363; NID:g4376677; PIDN:AAD18544.1; PID:g437668
 A:Experimental source: strain CWT029
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255
 A:Accession: H81584
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-257 <REA>
 A:Cross-references: GB:AE002198; GB:AE002161; NID:g7189279; PIDN:AAF38205.1; PID:g7189298
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CPn0400; CP0355
 C:Superfamily: Chlamydia trachomatis hypothetical protein CT254

Query Match 20.1%; Score 54; DB 2; Length 257;
Best Local Similarity 33.3%; Pred. No. 13;
Matches 16; Conservative 9; Mismatches 21; Indels

QY 2 FLHISSPFKYPTQEAQKEAQRSLGEMPGRHLGSSMSLALCLVPLVRE 49
|| : | |||| : |||| : | : : : | : |
Db 128 FLTLYLPTOEIHTEOVTQEVONSILPITGHYISMLNLGVLTPEGEE 173

RESULT 14

F86540
CT254 hypothetical protein [imported] - Chlamydophila pneumoniae (strain J138)
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: F86540
C:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: F86540
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <STO>
A:Cross-references: GB:BA000008; NID:98978772; PIDN:BA098608.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
C:Superfamily: Chlamydia trachomatis hypothetical protein CT254

Query Match	20.1%	Score 54;	DB 2;	Length 257
Best Local Similarity	33.3%;	Pred. No. 13;		
Matches 16; Conservative	9;	Mismatches 21;	Indels	
Qy	2	FLHTSSPKYPHTQEAQKQRSLGEMPGRHLSGSMISLALCLVPLVRE	49	
	:	: : : : : :		
Db	128	FLTVLVTQEILHTEQTVQESVNLS-PITGHYVISMLNLGV-LTFEGEE	173	

RESULT 15

B82990
Hypothetical protein PA5248 [Imported] - *Pseudomonas aeruginosa* (strain PAO1)
C/Species: *Pseudomonas aeruginosa*
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: B82990
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
-; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pa
A/Reference number: B82950; MUID:20437337
A/Accession: B82990
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-633 <STO>
A/Cross-references: GB:AE0040937; GB:AE004091; NID:g9951553; PIDN:AAG08633.1; GSPDB:GN
A/Experimental source: strain PAO1
C/Genetics:
A/Gene: PA5248

Query Match 20.1%; Score 54; DB 2; Length 633
Best Local Similarity 37.5%; Pred. No. 35;
Matches 18; Conservative 5; Mismatches 13; Indels

QY 12 PHTQEAK-----EAQRSLGEMPGRHLGSSMSLALCLVPLVREG 50
- - - - - : : : : :
Db 340 PETQVAOOLELAGKGLAEAAKQLG---GDSLSFSISFVSALLILREG 384

Search completed: July 30, 2001, 16:52:13
Job time: 248 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:48:05 ; Search time 26.85 seconds
(without alignments)
39.110 Million cell updates/sec

Title: US-09-402-713A-2

Perfect score: 268

Sequence: 1 MELHTSSPKYPTQEAQRE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	60.5	22.6	1239	2	US-08-937-931-2
2	60.5	22.6	1239	4	US-09-285-502-2
3	53	19.8	1151	1	US-08-286-889-37
4	53	19.8	1151	1	US-08-286-889-37
5	53	19.8	1151	1	US-08-485-618-37
6	53	19.8	1151	1	US-08-362-652-37
7	53	19.8	1151	2	US-08-605-672-37
8	53	19.8	1151	2	US-08-482-293A-37
9	53	19.8	1151	2	US-08-943-363-37
10	53	19.8	1151	4	US-09-193-043-37
11	53	19.8	1161	1	US-08-485-618-55
12	53	19.8	1161	1	US-08-362-652-55
13	53	19.8	1161	2	US-08-605-672-55
14	53	19.8	1161	2	US-08-482-293A-55
15	53	19.8	1161	2	US-08-943-363-55
16	52.5	19.6	1055	3	US-09-193-043-55
17	52.5	19.6	1055	3	US-09-031-563-21
18	52.5	19.6	1315	3	US-09-031-563-2
19	52.5	19.6	1315	3	US-09-031-563-25
20	51.5	19.2	269	1	US-08-447-534-5
21	51.5	19.2	269	1	US-08-468-763-17
22	51.5	19.2	269	1	US-08-448-160-5
23	51.5	19.2	269	2	US-08-393-996A-17
24	51	19.0	1155	1	US-08-286-889-46
25	51	19.0	1155	1	US-08-485-618-46
26	51	19.0	1155	1	US-08-362-652-46
27	51	19.0	1155	2	US-08-605-672-46

28	51	19.0	1155	2	US-08-482-293A-46	Sequence 46, Appl
29	51	19.0	1155	2	US-08-943-363-46	Sequence 46, Appl
30	51	19.0	1155	4	US-09-193-043-46	Sequence 46, Appl
31	51	19.0	1161	1	US-08-173-497-2	Sequence 2, Appli
32	51	19.0	1161	1	US-08-286-889-2	Sequence 2, Appli
33	51	19.0	1161	1	US-08-485-618-2	Sequence 2, Appli
34	51	19.0	1161	1	US-08-485-618-53	Sequence 53, Appl
35	51	19.0	1161	1	US-08-485-618-53	Sequence 99, Appl
36	51	19.0	1161	1	US-08-362-652-2	Sequence 2, Appli
37	51	19.0	1161	1	US-08-362-652-53	Sequence 2, Appli
38	51	19.0	1161	2	US-08-605-672-2	Sequence 2, Appli
39	51	19.0	1161	2	US-08-605-672-53	Sequence 53, Appl
40	51	19.0	1161	2	US-08-605-672-99	Sequence 99, Appl
41	51	19.0	1161	2	US-08-482-293A-2	Sequence 2, Appli
42	51	19.0	1161	2	US-08-482-293A-53	Sequence 53, Appl
43	51	19.0	1161	2	US-08-482-293A-99	Sequence 99, Appl
44	51	19.0	1161	2	US-08-943-363-2	Sequence 2, Appli
45	51	19.0	1161	2	US-08-943-363-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1
US-08-937-931-2
; Sequence 2, Application US/08937931
; Patent No. 5935792
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerald M.
; APPLICANT: Pan, DuoJia
; APPLICANT: Rooke, Jenny
; APPLICANT: Yavari, Reza
; APPLICANT: Xu, Tian
; TITLE OF INVENTION: KUZ: A No. 5935792a1 Family of Metalloproteases
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,931
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-937-931-2

Query Match 22.6%; Score 60.5; DB 2; Length 1239;
Best Local Similarity 47.2%; Pred. No. 2.9;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;
QY 9 FKYPHTQEAQKQKSLGEM----PGRHLGSSMSLA 40

```
Db 341 FKYPH-QKYTEAFAGAFYDSTGRRLLGSSANVA 375
||||| |: ||| |: | ||||| :|
;
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,889
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: P38,659
; REFERENCE/DOCKET NUMBER: 27866/32168
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-286-889-37

Query Match 19.8%; Score 53; DB 1; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 14 TQEAQKEAQRSLGPMGRHLGSSMSLALCLVPLVREG 50
||||| |: ||| |: | ||||| :|
Db 428 TQEARHWRPKS--EVNGTQIGSYFGASLCSDVDVDRG 462

RESULT 4
US-08-485-618-37
; Sequence 37, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,618
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: P38,659
; REFERENCE/DOCKET NUMBER: 27866/32168
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-286-889-37

Db 341 FKYPH-QKYTEAFAGAFYDSTGRRLLGSSANVA 375
||||| |: ||| |: | ||||| :|
;
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/285,502
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/937,931
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-285-502-2

Query Match 22.6%; Score 60.5; DB 4; Length 1239;
Best Local Similarity 47.2%; Pred. No. 2.9;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPHTQEAQKQARSLGEM---PGRHLGSSMSLA 40
||||| |: ||| |: | ||||| :|
Db 341 FKYPH-QKYTEAFAGAFYDSTGRRLLGSSANVA 375

RESULT 3
US-08-286-889-37
; Sequence 37, Application US/08286889
; Patent No. 5470953
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Mich
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
```



```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32797
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-485-618-37

Query Match 19.8%; Score 53; DB 1; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TQEAQKEAQRSLGEMPGRHLSMSLALCLVPLVREG 50
      ||||: :| | :| | :| | :| | :| | :| |
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462

RESULT 5
US-08-362-652-37
; Sequence 37, Application US/08362652
; Patent No. 5766850
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vlieten, Monica
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,652
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32391
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
US-08-362-652-37

Query Match 19.8%; Score 53; DB 1; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TQEAQKEAQRSLGEMPGRHLSMSLALCLVPLVREG 50
      ||||: :| | :| | :| | :| | :| | :| |
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462

RESULT 6
US-08-605-672-37
; Sequence 37, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vlieten, Monica
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605,672
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-605-672-37

Query Match 19.8%; Score 53; DB 2; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TQEAQKEAQRSLGEMPGRHLSMSLALCLVPLVREG 50
      ||||: :| | :| | :| | :| | :| | :| |
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462
```

```

STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-37

Query Match 19.8% Score 53; DB 2; Length 1151;
Best Local Similarity 37.88; Pred No 34;
Matches 14; Conservative 7; Mismatches 14; Indels

QY 14 TQEAQKAEQRLSGMPGRHLGSSMLALCLPLVLRG 50
||||: |:| |:| |:| |:| |:| |:| |:| |:|
Db 428 TQEARHWRPKS-EVRGTQIGSYFGASLCSVDVDRDG 462

RESULT 9
US-09-193-043-37
; Sequence 37, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193.043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Rattus rattus
; SS-09-193-043-37

```

```

Query Match      19.8%; Score 53; DB 4; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAKQAKRSLGEMPGHRLGSSMSIALCLVPLVREG 50
      ||||:  :| | | | | | | | | | | | | | | | | |
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSDVDVDRDG 462

RESULT 10
US-08-485-618-55
; Sequence 55, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485.618
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32797
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-618-55

Query Match      19.8%; Score 53; DB 1; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAKQAKRSLGEMPGHRLGSSMSIALCLVPLVREG 50
      ||||:  :| | | | | | | | | | | | | | | | | |
Db 438 TQEARHWRPKS--EVRGTQIGSYFGASLCSDVDVDRDG 472

RESULT 11
US-08-362-652-55
; Sequence 55, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; Sequence 55, Application US/08362652
; Patent No. 5766850
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362.652
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32391
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-362-652-55

Query Match      19.8%; Score 53; DB 1; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAKQAKRSLGEMPGHRLGSSMSIALCLVPLVREG 50
      ||||:  :| | | | | | | | | | | | | | | | | |
Db 438 TQEARHWRPKS--EVRGTQIGSYFGASLCSDVDVDRDG 472

RESULT 12
US-08-605-672-55
; Sequence 55, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```


;
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-943-363-55

Query Match 19.8%; Score 53; DB 2; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
||||: | | | | | : | | | | | : | | | | | : | |
Db 438 TOEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 472

RESULT 15
US-09-193-043-55
; Sequence 55, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: NO. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-193-043-55

Query Match 19.8%; Score 53; DB 4; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
||||: | | | | | : | | | | | : | | | | | : | |
Db 438 TOEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 472

Search completed: July 30, 2001, 16:50:50
Job time: 165 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:48:05 ; Search time 42.83 Seconds
(without alignments)
72.188 Million cell updates/sec

Title: US-09-402-713A-2

Perfect score: 268

Sequence: 1 MFLHSSPKFKYPTQEAQRE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	268	100.0	51	19 AAW79736	Prostate cancer an
2	268	100.0	51	19 AAW79738	Prostate cancer an
3	77	28.7	16	20 AAY07887	Human secreted pro
4	63	23.5	514	21 AAY87343	Human signal pepti
5	63	23.5	602	21 AAB43919	Human cancer assoc
6	60.5	22.6	1239	19 AAW56131	Drosophila melanoga
7	59	22.0	897	20 AAY07092	Colon cancer assoc
8	59	22.0	1141	21 AAB43008	Human ORF2772
9	56	20.9	11	19 AAW79737	Prostate cancer an
10	55	20.5	184	21 AAB40271	Human ORF35 p
11	55	20.5	2432	21 AAY85565	Human homologue of

12	54.5	20.3	99	20	AAW29863	Human secreted pro
13	54.5	20.3	100	20	AAW25713	Human secreted pro
14	54.5	20.3	461	17	AAR98341	Human p57 protein.
15	54.5	20.3	461	17	AAR98342	Bovine p57 protein
16	54.5	20.3	598	17	AAR98344	GST/truncated hui
17	54.5	20.3	673	17	AAR98343	GST/human p57 fusi
18	54	20.1	414	21	AAG22216	Arabidopsis thalia
19	54	20.1	414	21	AAG46290	Arabidopsis thalia
20	54	20.1	450	21	AAG22215	Arabidopsis thalia
21	54	20.1	450	21	AAG46289	Arabidopsis thalia
22	54	20.1	453	21	AAG22214	Arabidopsis thalia
23	54	20.1	453	21	AAG46288	Arabidopsis thalia
24	53.5	20.0	112	20	AAY60194	Human endometrium
25	53	19.8	1151	16	AAR78179	Rat alpha-d compos
26	53	19.8	1151	16	AAW23059	Rat beta 2 integri
27	53	19.8	1151	19	AAW72834	Rat alpha-d #2. R
28	53	19.8	1151	19	AAW5101	Rat beta-integrin
29	53	19.8	1151	19	AAW60001	Rat alpha d polype
30	53	19.8	1151	20	AAW73344	Rat alphas protein
31	53	19.8	1151	21	AAB07371	Rat alpha_d partia
32	53	19.8	1161	16	AAW78169	Rat alpha-d subuni
33	53	19.8	1161	18	AAW23062	Rat beta 2 integri
34	53	19.8	1161	19	AAW72824	Rat alpha-d #1. R
35	53	19.8	1161	19	AAW65104	Rat beta-integrin
36	53	19.8	1161	19	AAW60004	Rat alpha d polype
37	53	19.8	1161	20	AAW73345	Rat alphas protein
38	53	19.8	1161	21	AAB07374	Rat alpha_d protei
39	52.5	19.6	109	20	AAW48422	Human prostate can
40	52.5	19.6	1055	20	AAW31730	Human fused protei
41	52.5	19.6	1315	20	AAW43265	Human Fused Relate
42	52.5	19.6	1315	20	AAW31725	Human fused protei
43	52.5	19.6	1315	20	AAW31729	Human fused protei
44	52.5	19.6	1315	21	AAB30679	Amino acid sequenc
45	52.5	19.6	1333	20	AAW31727	Human fused cDNA i

ALIGNMENTS

RESULT 1
AAW79736
ID AAW79736 standard; Protein; 51 AA.
XX
AC AAW79736;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) protein variant 1.
XX
KW Prostate cancer antigen protein variant 1; PCA3; prostatic cancer;
PC.
XX
OS Homo sapiens.
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
DR N-PSDB; AAW62427.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
XX
PS Claim 16; Fig 2B-2J; 111pp; English.

XX The present sequence represents the prostate cancer antigen (PCA3).
 CC protein variant 1 encoded by a PCA3 spliced cDNA sequence comprising
 CC of exons 1, 2, 3, 4a and 4b of the PCA3 gene. The invention claims
 CC for PCA3 cDNA variants and the proteins they encode. The invention
 CC also claims for antibodies against PCA3 protein. The antibodies are
 CC claimed to be useful for detecting PCA3 protein in immunoassay tests,
 CC for diagnosing, assessing and prognosing of prostatic cancer (PC).
 CC Antibodies, optionally coupled to a cytotoxin or radioisotope, and
 CC nucleic acids antisense to PCA3 cDNA are claimed to be useful for
 CC treating PC, while determining elevated levels of PCA3 (as RNA or
 CC protein) is useful for detecting a predisposition to development of
 CC PC, e.g. in prenatal tests. Detecting PCA3 protein allows
 CC differentiation between malignant and benign prostatic disease,
 CC and the level of PCA3 expression allows correlation with the grade of
 CC tumour. PCA3 protein and its fragments are also claimed to be useful
 CC in vaccines for preventing PC; in drug screens for identifying
 CC specific (antagonists (potentially useful therapeutically) and for
 CC studying protein-DNA interactions.
 XX
 SQ Sequence 51 AA;

Query Match 100.0%; Score 268; DB 19; Length 51;
 Best Local Similarity 100.0%; Pred. NO. 4.7e-32;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLHSSPFKYPHTQEAQKRAQSLGEMPGRHGSSMSLALCLVPLVREGH 51
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 mflhsspfkyphtqeaqkraqslgempgrhlgssmslalcivplvreg 51

RESULT 2

AAW79738
 ID AAW79738 standard; Protein; 51 AA.

AC AAW79738;

XX 30-DEC-1998 (first entry)

XX Prostate cancer antigen (PCA3) wild-type protein.

XX Prostate cancer antigen protein variant 1; PCA3; prostatic cancer;
 KW PC.

XX Homo sapiens.

XX WO9845420-A1.

XX 15-OCT-1998.

XX 09-APR-1998; 98WO-CA00346.

XX 10-APR-1997; 97US-0041836.

XX (DIAG-) DIAGNOCURE INC.

XX Bussemakers MJG;

XX WPI; 1998-568347/48.

XX N-PSDB; AAV62430.

XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
 PT prevention and treatment of prostatic cancer

XX Claim 16; Fig 5B-5F; 111pp; English.

PS The present sequence represents the prostate cancer antigen (PCA3)
 XX wild-type protein sequence encoded by a PCA3 wild-type cDNA sequence
 CC comprising of exons 1, 2, 3, 4a-4d of the PCA3 gene. The invention
 CC claims for PCA3 cDNA variants and the proteins they encode. The
 CC invention also claims for antibodies against PCA3 protein. The
 CC antibodies are claimed to be useful for detecting PCA3 protein in

CC immunoassay tests, for diagnosing, assessing and prognosing of
 CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
 CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
 CC to be useful for treating PC, while determining elevated levels of
 CC PCA3 (as RNA or protein) is useful for detecting a predisposition to
 CC development of PC, e.g. in prenatal tests. Detecting PCA3 protein
 CC allows differentiation between malignant and benign prostatic disease,
 CC and the level of PCA3 expression allows correlation with the grade of
 CC tumour. PCA3 protein and its fragments are also claimed to be useful
 CC in vaccines for preventing PC; in drug screens for identifying
 CC specific (antagonists (potentially useful therapeutically) and for
 CC studying protein-DNA interactions.
 XX
 SQ Sequence 51 AA;

Query Match 100.0%; Score 268; DB 19; Length 51;
 Best Local Similarity 100.0%; Pred. NO. 4.7e-32;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLHSSPFKYPHTQEAQKRAQSLGEMPGRHGSSMSLALCLVPLVREGH 51
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 mflhsspfkyphtqeaqkraqslgempgrhlgssmslalcivplvreg 51

RESULT 3

AAV07887
 ID AAV07887 standard; Protein; 16 AA.

XX AAV07887;

XX 06-JUL-1999 (first entry)

XX Human secreted protein fragment encoded from gene 36.

XX Human; secreted protein; treatment; prevention; protein therapy; AIDS;
 KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
 KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
 KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
 KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
 KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
 KW arthritis; malignancy; digestive; endocrine; infection.

XX Homo sapiens.

XX WO9918208-A1.

XX 15-APR-1999.

XX 01-OCT-1998; 98WO-US20775.

XX 02-OCT-1997; 97US-0060884.

XX 02-OCT-1997; 97US-0060833.

XX 02-OCT-1997; 97US-0060836.

XX 02-OCT-1997; 97US-0060837.

XX 02-OCT-1997; 97US-0060838.

XX 02-OCT-1997; 97US-0060839.

XX 02-OCT-1997; 97US-0060843.

XX 02-OCT-1997; 97US-0060862.

XX 02-OCT-1997; 97US-0060866.

XX 02-OCT-1997; 97US-0060874.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;
 PI Florence KA, Greene JM, Janat F, Laflaur DW, NI J;
 PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
 XX WPI; 1999-264022/22.

XX N-PSDB; AAX37486.

XX New isolated human genes and the secreted polypeptides they encode


```
XX 08-MAR-2000; 2000WO-US050882.
PF
XX
XX
PR 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Rosen CA, Ruben SM;
PI
XX
XX WPI: 2000-587533/55.
XX
XX N-PSDB; AAC78128.
XX
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer
XX
XX
PS Claim 11; Page 2024-2026; 2352pp; English.
XX
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerable; immunomodulator;
CC antidiabetic; antilasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
XX the present invention.
XX
XX Sequence 602 AA;
SQ

Query Match 23.5%; Score 63; DB 21; Length 602;
Best Local Similarity 42.1%; Pred. No. 1.1;
Matches 16; Conservative 7; Mismatches 13; Indels 2; Gaps 2;

QY 12 PHTQEAQKQKSL-GEPMGRHIGSSMSLAL-CLVPLV 47
Db 559 praedlaedsgslygrapgrhtwslilaalacivp11 596
| : : : : | | | | | | : | | | | |
| : : : : | | | | | | : | | | | |

RESULT 6
AAW56131
ID AAW56131 standard; Protein; 1239 AA.
XX
XX
XX AAW56131;
XX
XX
XX 17-AUG-1998 (first entry)
XX
XX Drosophila melanogaster KUZ protein.
XX
XX kuzbanian; kuz; neurogenic; KUZ protein; neuronal partitioning;
XX development; NOTCH protein processing; regulation; cell function;
XX signal transduction pathways; screening; receptor binding;
XX metalloprotease.
XX
XX Drosophila melanogaster.
XX
XX WO9808933-A1.
XX
XX 05-MAR-1998.
XX
XX 27-AUG-1997; 97WO-US15099.
XX
```

```
XX 23-JUL-1997; 97US-0019390.
PR 29-AUG-1996; 96US-0019390.
XX
XX (REGC ) UNIV CALIFORNIA.
PA (UYIA ) UNIV YALE.
XX
XX Pan D, Rooke J, Rubin GM, Xu T, Yavari R;
PI
XX
XX WPI: 1998-179428/16.
XX
XX N-PSDB; AAV22647.
XX
XX
PT New KUZ polypeptides, members of the ADAM family of metalloprotease
PT - useful in neural partitioning and development
XX
XX
PS Claim 1; Pages 31-34; 58pp; English.
XX
XX
CC The sequence is that of encoded by the kuzbanian gene, a
CC neurogenic gene. The KUZ protein is involved in neuronal
CC partitioning and development. It is also involved in processing
CC of the NOTCH protein by cleaving the C-terminal portion. The
CC KUZ protein can be used to screen for compounds that alter
CC binding of KUZ to its receptor or the cleavage of the NOTCH
CC protein, hence acting to regulate NOTCH signal transduction
CC pathways and regulate cell functions.
XX
XX Sequence 1239 AA;
SQ

Query Match 22.6%; Score 60.5; DB 19; Length 1239;
Best Local Similarity 47.2%; Pred. No. 6.2;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 EKYPTQEAQKQKQKSLGEM----PGRHIGSSMSLA 40
Db 341 kfyph-qkytkeanfaegafydpstgrllgssanva 375
| | | | | | : | | | | | : |
| | | | | | : | | | | | : |

RESULT 7
AAV07092
ID AAV07092 standard; Protein; 897 AA.
XX
XX
XX AAV07092;
XX
XX
XX 02-JUL-1999 (first entry)
XX
XX
XX Colon cancer associated antigen precursor sequence.
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX prostate cancer.
XX
XX Homo sapiens.
XX
XX WO9904265-A2.
XX
XX
XX 28-JAN-1999.
XX
XX
XX 15-JUL-1998; 98WO-US14679.
XX
XX
XX 22-JUN-1998; 98US-0102322.
XX
XX 17-JUL-1997; 97US-0896164.
XX
XX 10-OCT-1997; 97US-0061599.
XX
XX 10-OCT-1997; 97US-0061765.
XX
XX 10-OCT-1997; 97US-0948705.
XX
XX 11-OCT-1997; 97GB-0021697.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
XX Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
XX Tureci O;
XX
```

DR WPI; 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products

PT for the diagnosis, monitoring or treatment of cancers

XX

PS Disclosure; Page 652-654; 787pp; English.

XX

CC The invention relates to a method for diagnosing a disorder characterised

CC by expression of a human cancer associated antigen precursor coded for by

CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a

CC biological sample isolated from a subject with an agent that specifically

CC binds to the NAM, an expression product or a fragment of an expression

CC product complexed with an HLA molecule; and (b) determining the

CC interaction between the agent and the NAM or the expression product as a

CC determination of the disorder. The products and methods can be used in

CC the diagnosis, monitoring, research, or treatment of conditions

CC characterised by the expression of various cancer associated antigens.

CC The invention provides nucleic acid sequences and encoded polypeptides

CC which are cancer associated antigen precursors expressed in human breast

CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

CC lung cancer.

XX

SQ Sequence 897 AA;

Query Match 22.0%; Score 59; DB 20; Length 897;

Best Local Similarity 32.0%; Pred. No. 6.9;

Matches 16; Conservative 14; Mismatches 16; Indels 4; Gaps 2;

QY 4 HISSPKYPHTQEAQKEAQRSL---GEMPGRHGSSMSLALCLVPLVREG 50

Db 198 hitaspkistqgeaerqalsirggtltgkfmsts-sipgcllgvaleg 246

RESULT 8

AAAB43008

ID AAB43008 standard; Protein; 1141 AA.

XX

AC AAB43008;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF2772 polypeptide sequence SEQ ID NO:5544.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;

KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;

KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;

KW thrombosis; contraceptive.

XX

OS Homo sapiens.

XX

PN WO200058473-A2.

XX

PD 05-OCT-2000.

XX

PF 31-MAR-2000; 2000WO-US08621.

XX

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Shimkets RA, Leach M;

XX

XX WPI; 2000-602362/57.

DR N-PSDB; AAC77217.

XX

PT Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,

PT neurodegenerative disorders and cardiovascular disease -

XX

PS Claim 11; Page 4728-4731; 5507pp; English.

XX

CC AAC774446 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnary;

CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

CC antidiabetic; hypotensive; dermatological; immunosuppressive;

CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;

CC antithyroid; and antianemic. The sequences can be used for determining

CC the presence of or predisposition to, or preventing or treating

CC pathological conditions associated with an ORFX-associated disorder. The

CC nucleic acids can be used to express ORFX proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,

CC graft vs host disease, cardiovascular disease, diabetes mellitus,

CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus

CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,

CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,

CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,

CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance

CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX

SQ Sequence 1141 AA;

Query Match 22.0%; Score 59; DB 21; Length 1141;

Best Local Similarity 32.0%; Pred. No. 9.3;

Matches 16; Conservative 14; Mismatches 16; Indels 4; Gaps 2;

QY 4 HISSPKYPHTQEAQKEAQRSL---GEMPGRHGSSMSLALCLVPLVREG 50

Db 405 hitaspkistqgeaerqalsirggtltgkfmsts-sipgcllgvaleg 453

RESULT 9

AAW79737

ID AAW79737 standard; Peptide; 11 AA.

XX

AC AAW79737;

XX

XX 30-DEC-1998 (first entry)

DT

XX

DE Prostate cancer antigen (PCA3) protein antigenic peptide.

XX

XX Prostate cancer antigen cDNA splice variant 1; PCA3; prostatic cancer;

KW PC; PCA3 antigenic peptide.

KW

XX

OS Homo sapiens.

XX

PN WO9845420-A1.

XX

PD 15-OCT-1998.

XX

PF 09-APR-1998; 98WO-CA00346.

XX

PR 10-APR-1997; 97US-0041836.

XX

PA (DIAG-) DIAGNOCURE INC.

XX

PI Bussemakers MJG;

XX WPI; 1998-568347/48.
 XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
 PT prevention and treatment of prostatic cancer
 XX
 XX Disclosure; Page 21; 11pp; English.
 XX
 CC The present sequence represents an antigenic peptide fragment
 CC derived from prostate cancer antigen (PCA3) protein. The invention
 CC claims for PCA3 cDNA variants and the proteins they encode. The
 CC invention also claims for antibodies against PCA3 protein. The
 CC antibodies are claimed to be useful for detecting PCA3 protein in
 CC immunosay tests, for diagnosing, assessing and prognosing of
 CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
 CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
 CC to be useful for treating PC, while determining elevated levels of
 CC PCA3 (as RNA or protein) is useful for detecting a predisposition
 CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
 CC allows differentiation between malignant and benign prostatic disease,
 CC and the level of PCA3 expression allows correlation with the grade of
 CC tumour. PCA3 protein and its fragments are also claimed to be useful
 CC in vaccines for preventing PC; in drug screens for identifying
 CC specific (ant)agonists (potentially useful therapeutically) and for
 CC studying protein-DNA interactions.
 XX
 SQ Sequence 11 AA;
 Query Match 20.9%; Score 56; DB 19; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.093;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 HTOEAOKEAQR 23
 Db | | | | | | | | | |
 1 htgaqkeagr 11
 RESULT 10
 AAB40271
 ID AAB40271 standard; Protein; 184 AA.
 XX
 AC AAB40271;
 XX
 XX 08-FEB-2001 (first entry)
 XX Human ORFX ORF35 polypeptide sequence SEQ ID NO:70.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-US08621.
 PF
 XX 31-MAR-1999; 99US-0127607.
 PR
 PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.
 XX 30-MAR-2000; 2000US-0540763.
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR N-PSDB; AAC74480.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 455; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 184 AA;
 Query Match 20.5%; Score 55; DB 21; Length 184;
 Best Local Similarity 34.0%; Pred. No. 3.9;
 Matches 16; Conservative 3; Mismatches 16; Indels 12; Gaps 1;
 QY 6 SSPKYPHTQEAQKEAQRSLG-----EMPCRHLGSSMSLA 40
 Db | | | | | | | | | | : | | | | | | | |
 77 smpgkspapapekegersrsrsklssglpqkqpqldgrhsssssia 123
 RESULT 11
 AAY85565
 ID AAY85565 standard; Protein; 2432 AA.
 XX
 AC AAY85565;
 XX
 XX 07-JUL-2000 (first entry)
 XX Human homologue of UNC-53 (Hs-UNC-53/2) sequence.
 DE
 XX UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
 KW anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
 KW antisclerotic; antimetastatic; anti-arthritis; autoimmune disease.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Region 1..89
 FT /note= "this region can be replaced with one of the
 FT three sequences shown in AAY85566 to AAY85568;
 FT this creates three variants at the N-terminus"
 FT Misc-difference 1018
 FT /label= Asp or Glu

```

Region 1776..1778
/note= "present or absent depending upon the allele
from which the protein is translated"
WO9963080-A1.
09-DEC-1999.
02-JUN-1999; 99WO-EP03848.
03-JUN-1998; 98GB-0011962.
(JANC ) JANSSEN PHARM NV.
Luyten WHML, De Raeymaeker MC, Geysen JJGH, Bogaert TAOE;
Maerten LJS, Verhasselt P, Van De Craen M;
WPI: 2000-116370/10.
N-PSDB: AAA07836.
Novel proteins and nucleic acids e.g. for treating neurodegeneration -
Claim 95; Fig 1d; 146pp; English.
The invention provides vertebrate (human) protein homologue of a UNC-53
protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
their plus ends. The UNC-53 sequences are used to promote neural
regeneration, revascularization and wound healing; also for treating
neurodegenerative disease, acute traumatic injury, fibrotic disease and
autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
polynucleotides can be used for recombinant production of the proteins,
as a source of probes for detecting allelic variants and polymorphisms,
for sequencing genomic DNA and for detecting UNC-53 expression; and as
source of therapeutic antisense sequences. Cells that express the
protein are used to identify regulators of cell shape, growth, motility
and migration. They can also be used to identify proteins that are
involved in signal transduction pathways also involving UNC-53, and to
identify compounds that alter attachment of UNC-53 to microtubules. A
target gene coupled to a UNC-53 encoding sequence may be used to deliver
the target gene to a cellular microtubule or its plus ends. The present
sequence represents the amino acid sequence of the second human homologue
of UNC-53, designated hs-UNC-53/2.
Sequence 2432 AA;
Query Match 20.58; Score 55; DB 21; Length 2432;
Best Local Similarity 34.08; Pred. No. 90;
Matches 16; Conservative 3; Mismatches 16; Indels 12; Gaps 1
QY 6 SSPEKYPTQEAQKQAEQSLG-----EMPGRHLSGMSLA 40
| | | | | | | | | | | | | | | | | | | | | |
Db 571 smpkpsapapskegersrgklssgipqkqpdidgrhsssssla 617
| | | | | | | | | | | | | | | | | | | | | |
RESULT 12
AAAY29863
ID AAY29863 standard; Protein; 99 AA.
XX
AC AAY29863;
XX
17-NOV-1999 (first entry)
XX
Human secreted protein clone gn8_6.
XX
Human; secreted protein; biological activity; nutritional; cytokine;
KW cell proliferation; differentiation; immune stimulating; vaccine;
KW hematopoiesis regulation; tissue growth; haemostatic; thrombolytic;
KW anti-inflammatory; tumour inhibition.
XX
OS Homo sapiens.
XX
PN WO9946287-A1.

```


FT Peptide /note= "WD40 repeat fragment"
FT 127..161
FT Peptide /note= "WD40 repeat fragment"
FT 172..203
FT Peptide /note= "WD40 repeat fragment"
FT 215..252
FT Peptide /note= "WD40 repeat fragment"
FT 265..298
FT Peptide /note= "WD40 repeat fragment"
FT 433..461
FT Peptide /note= "Leucine rich C terminal fragment"

XX JP08111996-A.
XX 14-MAY-1996.
XX 21-OCT-1994; 94JP-0282743.
XX 21-OCT-1994; 94JP-0282743.
XX (NISB) JAPAN TOBACCO INC.
XX WPI; 1996-283507/29.
XX N-PSDB; AAT30361.
XX A new protein, p57, comprising WD40 repeat region - used for
XX development of anti-cancer and anti-HIV agents
XX Example 2; Page 27-29; 5lpp; Japanese.

XX This sequence represents the bovine p57 protein. p57 contains a leucine
XX rich C-terminal peptide which comprises a leucine residue after each
XX 7 amino acids and a WD40 repeated structure region containing five
XX WD40 regions. p57 forms a dimer. It can be used in the development
XX of an anti-cancer agent and an anti-HIV agent. It can also combine
XX with actin in the control of cell movement.

SQ Sequence 461 AA;

Query Match 20.3%; Score 54.5; DB 17; Length 461;
Best Local Similarity 43.8%; Pred. No. 14;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

OY 2 FLHSSPFKYPHTQEAQKQPSLGLGEMPCRH 33
| | | | : | | | | |
Db 303 flhysmf-----skesqrgmgypkrgl 327

Search completed: July 30, 2001, 16:51:39
Job time: 214 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 04:11:36 ; Search time 6114.61 Seconds
(without alignments)
2894.010 Million cell updates/sec

Title: US-09-402-713A-3
Perfect score: 1872
Sequence: 1 agaagctggcatcagaaaaa.....caataaagaattacaaga 1872

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1028115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: gb_estl1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_est4:*
- 5: gb_est5:*
- 6: gb_est6:*
- 7: gb_est7:*
- 8: gb_est8:*
- 9: gb_est9:*
- 10: gb_est10:*
- 11: gb_est11:*
- 12: gb_est12:*
- 13: gb_est13:*
- 14: gb_est14:*
- 15: gb_est15:*
- 16: gb_est16:*
- 17: gb_est17:*
- 18: gb_est18:*
- 19: gb_est19:*
- 20: gb_est20:*
- 21: gb_est21:*
- 22: gb_est22:*
- 23: gb_est23:*
- 24: gb_est24:*
- 25: gb_est25:*
- 26: gb_est26:*
- 27: gb_est27:*
- 28: gb_est28:*
- 29: gb_est29:*
- 30: gb_est30:*
- 31: gb_est31:*
- 32: gb_est32:*
- 33: gb_est33:*
- 34: gb_est34:*
- 35: gb_est35:*
- 36: gb_est36:*
- 37: gb_est37:*
- 38: gb_est38:*
- 39: gb_est39:*
- 40: gb_est40:*
- 41: gb_est41:*
- 42: gb_est42:*
- 43: gb_est43:*
- 44: gb_est44:*
- 45: gb_est45:*
- 46: gb_est46:*
- 47: gb_est47:*

- 44: em_esthum10:*
- 45: em_esthum11:*
- 46: em_esthum12:*
- 47: em_esthum13:*
- 48: em_esthum14:*
- 49: em_esthum15:*
- 50: em_esthum16:*
- 51: em_esthum17:*
- 52: em_esthum18:*
- 53: em_esthum19:*
- 54: em_esthum20:*
- 55: em_esthum21:*
- 56: em_esthum22:*
- 57: em_esthum23:*
- 58: em_esthum24:*
- 59: em_esthum25:*
- 60: em_esthum26:*
- 61: em_esthum27:*
- 62: em_esthum28:*
- 63: em_estin1:*
- 64: em_estin2:*
- 65: em_estin3:*
- 66: em_estin4:*
- 67: em_estin5:*
- 68: em_estom1:*
- 69: em_estov1:*
- 70: em_estov2:*
- 71: em_estpl1:*
- 72: em_estpl2:*
- 73: em_estpl3:*
- 74: em_estpl4:*
- 75: em_estpl5:*
- 76: em_estpl6:*
- 77: em_estpl7:*
- 78: em_estpl8:*
- 79: em_estpl9:*
- 80: em_estpl10:*
- 81: em_estro1:*
- 82: em_estro2:*
- 83: em_estro3:*
- 84: em_estro4:*
- 85: em_estro5:*
- 86: em_estro6:*
- 87: em_estro7:*
- 88: em_estro8:*
- 89: em_estro9:*
- 90: em_estro10:*
- 91: em_estro11:*
- 92: em_estro12:*
- 93: em_estro13:*
- 94: em_estro14:*
- 95: em_estro15:*
- 96: em_estro16:*
- 97: em_estro17:*
- 98: em_estro18:*
- 99: em_estro19:*
- 100: em_estro20:*
- 101: gb_est25:*
- 102: gb_est26:*
- 103: gb_est27:*
- 104: gb_est28:*
- 105: gb_est29:*
- 106: gb_est30:*
- 107: gb_est31:*
- 108: gb_est32:*
- 109: gb_est33:*
- 110: gb_est34:*
- 111: gb_est35:*
- 112: gb_est36:*
- 113: gb_est37:*
- 114: gb_est38:*
- 115: gb_est39:*
- 116: gb_est40:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

COMMENT

Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
Seq primer: 77
Class: BAC ends.

FEATURES

Location/Qualifiers
1..654
/organism="Homo sapiens"
/db_xref="GDB:754135"
/db_xref="taxon:9606"
/clone="RPCI-11-108L4"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT 213 a 100 c 139 g 202 t
ORIGIN

Query Match 32.8%; Score 613.8; DB 227; Length 654;
Best Local Similarity 98.5%; Pred. No. 4.4e-152;
Matches 645; Conservative 4; Mismatches 3; Indels 3; Gaps 3;

QY 1030 tccatttcgaagcctcaaatgtcattccatttaataatcacagagattatctttttt 1089
DB 654 TCCCATTTCTCAAGACCTCAAAATGTCAATCCATTAATATCACAGGATTAACATTTT 595
QY 1090 ttaacctggaagaattcaatgttaccatgcagctatgggaatttaattacattttgttt 1149
DB 594 TTAACCTGGAAGAATTCAATGTACATGCAGCTATGGGAATTAATACATATTTTGT 535
QY 1150 tccagtgaagaatgactaagtccttttccctccctttgtttgttttttttcagta 1209
DB 534 TCCAGTGAAGAATGACTAAGTCTTTATCCCTCCCTTTGTGATTTTTCAGTA 475
QY 1210 taaagttaaaatgcttagccttgcctgaggtgtatatacag-cacagcctctcccatcc 1268
DB 474 TAAAGTAAATGCTTAGCCTTGTACTGAGGCTGTATACAGCCACAGCCTCTCCCATCC 415
QY 1269 ctccagccttctgtcatcacatcaacccctcccatnysacctaaacaaatctaaact 1328
DB 414 CTCCAGCCTTATCTGTCACTACCATCAACCCCTCCCATG-CACCTAAACAAATCTA 356
QY 1329 tgaattctcttgaacatgtcaggnatcacattttctctctgctgagaagctcttccctt 1388
DB 355 TGTATTTCTTGAACATGTGAGG-CATACATATTCTTCTGCTGAGAAGCTCTTCTCT 297
QY 1389 gctcttaantctagaatgatgtataagtttgaatgaagtgtactatcttacttcagtaa 1448
DB 296 GTCTCTTAATAGTAAATGATGTAAAGTTTGAATTAAGTTGACTATCTTACTTCTATG 237
QY 1449 aagaaggacacatatgagattcatcatcacatgagacagacaaatctaaagtgttaattt 1508
DB 236 AGAAGGACACATATGAGATTTCATCATCATGAGACACAAATACTAAAGTGTAAATTT 177
QY 1509 gattataagagtttagataaaatataatgaatgcaagkccacagaggggaagtgtttatgg 1568
DB 176 GATTATAAGAGTTAGATAAATATATGAATCAAGAGCCACAGAGGGAATGTTTATGG 117
QY 1569 gcaagtgttgagcctcggtgatgtgaagmaaagcgaggaacctcatagttattataaa 1628
DB 116 GCACGTTTCTAAGCCTGGGATGTGAAGCAAGGACGGAACCTCATAGTATCTTATATAA 57

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	613.8	27.4	654	227	AQ319247
C 2	512.2	32.8	856	21	A1557225
C 3	399.6	21.3	441	170	BF858890
C 4	394	21.0	402	9	AA578773
C 5	362	19.3	394	225	AQ206972
C 6	306.2	16.4	332	170	BF858286
C 7	260.4	13.9	290	147	BF373619
C 8	226.8	12.1	282	147	BF373581
C 9	214.4	11.5	226	170	BF858371
C 10	167	8.9	167	147	BF373406
C 11	138.2	7.4	451	226	AQ246715
C 12	86.8	4.6	657	21	A1557495
C 13	63	3.4	1101	219	CNS0039G
C 14	56	3.0	1001	219	CNS01400
C 15	52.6	2.8	1101	219	CNS0182P
C 16	52.2	2.8	1101	219	CNS0100X
C 17	51.6	2.8	1101	219	CNS000B1
C 18	49.2	2.6	1204	219	CNS016E2
C 19	48.2	2.6	1101	219	CNS017KX
C 20	47.8	2.6	1101	219	CNS0039G
C 21	47.6	2.5	529	225	AQ173559
C 22	47.6	2.5	1101	219	CNS0039E
C 23	47.6	2.5	1101	219	CNS00EVL
C 24	47.6	2.5	1101	219	CNS010EY
C 25	47.2	2.5	1201	219	CNS0160R
C 26	46.8	2.5	654	223	AQ046642
C 27	45.2	2.4	1086	219	CNS00YXK
C 28	44.8	2.4	938	219	CNS006YI
C 29	44.8	2.4	943	219	CNS002IM
C 30	44.8	2.4	1125	106	AL547503
C 31	44.4	2.4	926	219	CNS008LM
C 32	44.4	2.4	1101	219	CNS017KT
C 33	43.6	2.4	1101	219	CNS00L72
C 34	43.2	2.3	352	12	AA791841
C 35	43.2	2.3	656	119	AW691330
C 36	43.2	2.3	997	219	CNS0035E
C 37	43.2	2.3	1101	219	CNS017KE
C 38	43	2.3	928	219	CNS0076D
C 39	43	2.3	930	219	CNS00D23
C 40	42.8	2.3	821	142	BE960968
C 41	42.8	2.3	1101	219	CNS0039L
C 42	42.6	2.3	859	219	CNS004YY
C 43	42.6	2.3	1077	222	CNS05AA9
C 44	42.4	2.3	734	219	CNS010MP
C 45	42.4	2.3	1092	220	CNS020K7

ALIGNMENTS

RESULT 1
AQ319247/c
LOCUS AQ319247 654 bp DNA GSS 06-MAY-1999
DEFINITION RPC111-108L4..TV Human sapiens genomic clone RPCI-11-108L4,
DNA sequence.
ACCESSION AQ319247
VERSION AQ319247.1 GI:4052212
KEYWORDS GSS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 654)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P., and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)

Qy 1128 aatttaataacataattttgttttcagtgcaagatgactaagtctcttattatccctccct 1187
|||||
Db 441 AATTTAATACATAATTTGTTTTCAGTGCAGAGATGACTAGTCTCTTATCCCTCCCT 382
Qy 1188 ttgtttgattttttccagataaagttaaaatccttagcttgactgaggtgtata 1247
|||||
Db 381 TTGTTTGATTTTTCAGTATAAGTTAAATGCTTAGCTTACGTAGGCTGTATA 322
Qy 1248 cag-cacagctctcccatccctccagcttattgtcattccatcaacccctcccat 1306
|||||
Db 321 CAGCCACAGCCCTCTCCCATCCCTCCAGCTTATCTGTCATCACCATCAACCCCTCCCAT 262
Qy 1307 nysacttaaaacaaatctaaattgtaattccttgcaactgtcaggncatatacttrtct 1366
|||||
Db 261 G-CACCTAAACAAATCTAACTTGTAATCTTGACATGTCAGG-CATACATTAATCTCT 204
Qy 1367 ttgctgagagct 1426
|||||
Db 203 TCTGCTGAGAGCT 144
Qy 1427 ttgactatcttacttcagcaagaagagagacacatatgagattcattcattcattcatt 1486
|||||
Db 143 TTGACTATCTTACTTCATGCAAGAAGGACACATATGAGATTTCATCATCATGAGACA 84
Qy 1487 gcaataactaaaagttaattgatttaaggttttagataaaatatgaaatgcaagak 1546
|||||
Db 83 GCATAATACCAAGTGTAATTTGATTATTAAGAGTTTAGATAAATATATGAATGCAAGAG 24
Qy 1547 ccacagaggggaattgtttggg 1568
|||||
Db 23 CCACAGAGGGAATGTTTATGGG 2

RESULT 4
AA578773 402 bp mRNA EST 12-SEP-1997
LOCUS nh24a04.s1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:953262, mRNA
DEFINITION sequence.
ACCESSION AA578773
VERSION AA578773.1 GI:2356957
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.,
Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 565 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 373.
Location/Qualifiers
1. .402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:953262"
/clone.lib="NCI_CGAP_Prl"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"

/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected, histologically normal prostate epithelial cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the uDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

BASE COUNT 128 a 86 c 84 g 104 t
ORIGIN

Query Match 21.0%; Score 394; DB 9; Length 402;
Best Local Similarity 98.8%; Pred. No. 9.4e-94;
Matches 397; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 395 ggatgggcaggaacacagatcctgttggtatattttgaacgggattacagattga 454
|||||
Db 1 GGTCCACAGGAAACAGATCCTGTGTGGATATATTTGAACGGGATTACAGATTGA 60
Qy 455 aatgaagtcacaaatgtgagcattaccatgagagaaacacagagaaaaatccttgatgg 514
|||||
Db 61 AATGAAGTCACAAAGTGAGCATTACCAATGAGAGGAAACACAGAGAAATCTTGTATGG 120
Qy 515 ctccacagacatgac 574
|||||
Db 121 CTTCAAGACATGCAACAAACAAATGGAATGCTGTATGATGATGAGGACGCCAAGCTG 180
Qy 575 gggagagataaccacagggcagaggtcagattctgacctgctgacctgctgacctgctg 634
|||||
Db 181 GGGAGAGATAACACAGGGCAGGGTCAGGATTTCTGCCCTCTGCTTAACCTGTGGC 240
Qy 635 ttcaataccaaatcatcttcatatttcaacctcaaaacacacacacacacacacacacac 694
|||||
Db 241 TTCAATACCAAAATCATTTTCATATTTCTAACCCCTCAAAACAAAGCTGTCTTAATATCTGAT 300
Qy 695 ctctcaggttctctctctggggccacacattctccatatatccacgacacactcttttaata 754
|||||
Db 301 CTCTACGGTTCTCTTGGGCCCAACATTTCTCATATATATCCAGCCACACTCATTTTAAATA 360
Qy 755 tttagtccagatctgtactgtgacctttctacactgtaga 796
|||||
Db 361 TTTAGTCCAGATCTGTACTGTGACCTTCTACACTGTAGA 402

RESULT 5
AQ206972/3 394 bp DNA GSS 17-SEP-1998
LOCUS HS_3238_B1_G11_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3238 Col=21 Row=N, DNA sequence.
ACCESSION AQ206972
VERSION AQ206972.1 GI:3617542
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 394)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618

AUTHORS
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL MEDLINE COMMENT
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=MR0st2-MR0-FT0175-310800-106-h09st3-2000-08-31st4-1)
Seq primer: puc 18 forward
High quality sequence stop: 290.
Location/Qualifiers
1. .290
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0175"
/dev_stage="Adult"

FEATURES
source
1. .290
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0175"
/dev_stage="Adult"
/site="Organ: prostate_tumor; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
78 a 73 c 53 g 86 t

BASE COUNT
78 a 73 c 53 g 86 t

ORIGIN

Query Match 13.9%; Score 260.4; DB 147; Length 290;
Best Local Similarity 97.8%; Pred. No. 2.7e-58;
Matches 264; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 575 gggaggagataaacacggggcagagggtcaggattctggccctgctgctaaactgtgcg 634
Db 21 gggaggagataaacacggggcagagggtcaggattctggccctgctgctaaactgtgcg 80
Qy 635 ttcataacccaatcatttcatttctaacctcaaaacaaagctgtgtgtaatactgat 694
Db 81 TTCATAACCAATCATTTTCATATTTCTAACCTCAAAACAAAGCTGTGTAATATCTGAT 140
Qy 695 ctctacggttccctctggggcccaacattctccatatatccagccacacatcttttaata 754
Db 141 CTCTACGGTTCCTCTGGGGCCCAACATCTCCATATATCCAGCCACACTCAATTTTAATA 200
Qy 755 tttaagttccagatctgactgtgacctcttcacacttagaataaacattactctatttg 814
Db 201 TTTAGTTCCTCCAGATCTGACTGTGACCTTTCTACACTGTAGAATAACATTACTCAATTTGG 260
Qy 815 ttcaagacccttcgtgtgtgcctaata 844
Db 261 TTCAGACCCCTTCGTGTGGCTAACGACTA 290

RESULT
8
BF373581/c 282 bp mRNA EST 24-NOV-2000
LOCUS BF373581
DEFINITION MR0-FT0175-210800-101-d05 FT0175 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF373581
VERSION BF373581.1 GI:11335606
KEYWORDS EST.

SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 282)

REFERENCE
AUTHORS
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=MR0st2-MR0-FT0175-210800-101-d05st3-2000-08-21st4-1)
Seq primer: puc 18 forward
High quality sequence stop: 282.
Location/Qualifiers
1. .282
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0175"
/dev_stage="Adult"
/site="Organ: prostate_tumor; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
85 a 51 c 70 g 76 t

BASE COUNT
85 a 51 c 70 g 76 t

ORIGIN

Query Match 12.1%; Score 226.8; DB 147; Length 282;
Best Local Similarity 91.3%; Pred. No. 2.2e-49;
Matches 252; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

Qy 568 caagctggggaggagataaacacggggcagagggtcaggattctggccctgctgctaaa 627
Db 282 CAAGCTGGGAGGAGATAAACACGGGGCAGAGGGTCAGGATTCTGGCCCTGCTGCTAAA 223
Qy 628 ctgtcgctcataacaaaatcatttcatttcaacctcaaaacaaagctgtgtgtaat 687
Db 222 CTGTGCGTTTCATAACCAATCATTTTCATATTTCTAACCTCAAAACAAAGCTGTGTAAT 163
Qy 688 atctgatctctacggttccctctggggcccaacattctccatatatccagccacacatt 747
Db 162 ATCTGATCTACGGTTCTCTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCAT 103
Qy 748 tttaataatttagttccagatctgactgtgaccttcttcacacttagaataacatta 805
Db 102 TTTAATATTTAGTTCCTCCAGATCTGACTGTGACCTTTCTTCTACATCTGTAGAATAACATTA 43
Qy 806 ctctattgttcaagacccttcgtgtgtgccta 841
Db 42 CTCATTTGTTCAAGACCTTCCTACTTGTAGCGACTA 7

RESULT
9

DEFINITION HS_2059_B2_E08_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2059 Col=16 Row=J, DNA sequence.

ACCESSION A0246715

VERSION A0246715.1 GI:3696897

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Sequence Tagged Connector
Plate: 2059 row: J column: 16
Class: BAC ends
High quality sequence stop: 451.

FEATURES
source
1. .451
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2059 Col=16 Row=J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 133 a 101 c 86 g 130 t 1 others

ORIGIN

Query Match 7.4%; Score 138.2; DB 226; Length 451;
Best Local Similarity 94.1%; Pred. No. 8.8e-26;
Matches 143; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1721 tcattgtgaagtcgcttagaatttgcaatcactactgtgtcacttctcactttga 1780
|||||
Db 2 TCATGGTGAGTCGCTTTAGAAATGTGGCAATCATCTGGTCACCTATCTCAACTTTGA 61
|||||

Qy 1781 gatgttcttccttgaatttgaaagaataggcactcttctgagccactttagg 1840
|||||
Db 62 GATGTGTGCGCTTGTAGTGNATTCAGAAATACGGCACCTCTTGTGAGCCACTTTAGG 121
|||||

Qy 1841 gttcactctggcaataagaatttacaaga 1872
|||||
Db 132 GTCACCTCTGGCAATAAAGAAATACANAGA 153
|||||

RESULT 12

AI557495 657 bp mRNA EST 09-AUG-1999

DEFINITION PT2.1.7_H12.r tumor2 Homo sapiens CDNA 3', mRNA sequence.

ACCESSION AI557495

VERSION AI557495.1 GI:4489858

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Huang,G.M., Ng,W.I., Farkas,J., He,L., Liang,H.A., Gordon,D., Yu,J. and Hood,L.

TITLE Prostate cancer expression profiling by cDNA sequencing analysis

JOURNAL Genomics 59 (2), 178-186 (1999)

MEDLINE 99339982

COMMENT Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huangm@yahoo.com.

FEATURES
source
1. .657
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."

BASE COUNT 179 a 112 c 153 g 166 t 47 others

ORIGIN

Query Match 4.6%; Score 86.8; DB 21; Length 657;
Best Local Similarity 97.8%; Pred. No. 4.4e-12;
Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 11 atcagaaaaacagagggagatttctgtgctgcagccagggagaccaggaagatctgc 70
|||||
Db 12 AGCAGAAAAACAGAGGGGAGATTTGTGTGGCTGACGCCGAGGACCAAGGAATCTGC 71
|||||

Qy 71 atggtgggaagacctgatgacagaggt 100
|||||
Db 72 ATGTGGGAAGGACCTGATGATACAGAGTT 101
|||||

RESULT 13

CNS0039G 1101 bp DNA GSS 03-JUN-1999

LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #

DEFINITION BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL063921

VERSION AL063921.1 GI:4941778

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

AUTHORS 1 (bases 1 to 1101)

TITLE Genoscope.

JOURNAL Direct Submission

COMMENT Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see [http://www.fruitfly.org/TheBDGP/Drosophila melanogaster BAC library](http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library) was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1. .1101
Location/Qualifiers

COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES Location/Qualifiers
source 1..1101
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN37D10"
/note="end : SP6"
BASE COUNT 274 a 268 c 128 g 73 t 358 others
ORIGIN

Query Match 2.8%; Score 52.6; DB 219; Length 1101;
Best Local Similarity 19.5%; Pred. No. 0.0061;
Matches 104; Conservative 196; Mismatches 233; Indels 0; Gaps 0;
QY 620 tgcctaaactgtcgctcatacacaatcttctcattcttaacccctcaaaacaagct 679
Db 1091 THMAHHCHCMHYMMHYMHYMHYHCHYMAAHAYMHHTAYAHMCHAHHTWTHYWT 1032
QY 680 gtgttaatactgtctcactcgttctctctcgtgcccacattctccatataccagcca 739
Db 1031 CWYMYAWYCVCYCHHWHWHYAWYMMWHYAWMHWHHHHMMCMCMHYMMHMCYMC 972
QY 740 cactcatttttaatttccagatctcgtactgtgaccttctcactcgttagaata 799
Db 971 YCCWHMWHCHHHTATTHHHYTHYMMWTCHTWTHWYHMTMHHWHHHAAMAHA 912
QY 800 acattactcatttgttcaagaccctcgtgtgctgctaatatgtagtactgttt 859
Db 911 ATHTWYHMAHHTYTYTYTYTYTYTYTYTYTYTYTYTHHWTAAHHHWHYAHWHHWH 852
QY 860 ttctaaaggagtgtctgtgccagggatctgtgaacagggtgggaagcatctcaagatc 919
Db 851 WYHYWAAWTTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHTTTH 792
QY 920 ttccagggttactactagcacagcatgattcattacgaggtgaattatctaata 979
Db 791 CCHYTANYTHMYCTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHT 732
QY 980 acatcatcctcagtgcttgcctcactgaaatcattccacttttgcctcattct 1039
Db 731 CCHMCHCMHYMMYMCCTHTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHT 672
QY 1040 caagacctcaaatgtcatctcattataatcacaggttaacttttttttaacctgga 1099
Db 671 HTTTCCHCHTTCMCAMCWCCTTTTAAACACTTTTTHHHTTTHHHTTTHHHTTTHH 612
QY 1100 agaattcaatgtacatgcagctatgggaatttaattacatatttggtttcc 1152
Db 611 AWWTTWTTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHTTTHHHT 559

Search completed: July 31, 2001, 04:11:46
Job time: 9998 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:18:22 ; Search time 176.8 seconds
(without alignments)
2004.471 Million cell updates/sec

Title: US-09-402-713A-3

Perfect score: 1872

Sequence: 1 agaagctggcatcagaaaaa.....caataaagaatttacaaga 1872

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59.8	3.2	7218	1	US-08-232-463-14
2	57.4	3.1	7218	1	US-08-232-463-14
3	47	2.5	10968	2	US-08-680-327-2
4	47	2.5	10968	4	US-09-228-246-1
5	45.4	2.4	5134	2	US-08-310-912A-157
6	45.4	2.4	5134	5	PCT-US95-04589-157
7	45.4	2.4	5475	2	US-08-680-327-1
8	45.4	2.4	5475	4	US-09-228-246-3
9	35.8	1.9	14507	3	US-08-785-150-1
10	35	1.9	4129	2	US-08-370-319C-12
11	35	1.9	4129	4	US-09-224-834-12
12	34.8	1.9	5718	3	US-08-714-918-48
13	34.8	1.9	5718	4	US-09-265-315-48
14	34.8	1.9	5718	4	US-09-265-315-48
15	34.8	1.9	5718	4	US-09-266-417-48
16	34.2	1.8	2679	1	US-07-977-434-11
17	34.2	1.8	2679	1	US-08-458-819-11
18	34.2	1.8	2679	5	PCT-US91-07035-11
19	33.8	1.8	2570	2	US-09-056-075-2
20	33.8	1.8	246240	2	US-08-724-394A-20
21	33.8	1.8	246240	2	US-08-724-394A-21
22	33.8	1.8	246240	2	US-08-724-394A-22
23	33.6	1.8	3084	1	US-08-551-437-3
24	33.6	1.8	3084	3	US-09-004-225-3
25	33.6	1.8	2084	4	US-09-084-346-3
26	33.2	1.8	5173	1	US-08-242-677-1
27	33	1.8	1952	1	US-08-333-358-1

28	33	1.8	1952	1	US-08-463-694-1	Sequence 1, Appli
29	33	1.8	1952	1	US-08-694-501-1	Sequence 1, Appli
30	33	1.8	2950	5	PCT-US93-08386-7	Sequence 7, Appli
31	33	1.8	2968	5	PCT-US93-08386-1	Sequence 1, Appli
32	33	1.8	8920	2	US-08-446-855A-1	Sequence 1, Appli
33	33	1.8	8920	4	US-09-150-741-1	Sequence 1, Appli
34	32.8	1.8	720	4	US-08-446-935-4	Sequence 4, Appli
35	32.8	1.8	2395	4	US-08-446-935-7	Sequence 7, Appli
36	32.8	1.8	2405	1	US-08-454-097-30	Sequence 30, Appli
37	32.8	1.8	2405	3	US-08-185-359-30	Sequence 30, Appli
38	32.8	1.8	5356	4	US-08-446-935-1	Sequence 1, Appli
39	32.6	1.7	3607	1	US-08-647-351B-1	Sequence 1, Appli
40	32.6	1.7	246240	2	US-08-724-394A-20	Sequence 20, Appli
41	32.6	1.7	246240	2	US-08-724-394A-21	Sequence 21, Appli
42	32.6	1.7	246240	2	US-08-724-394A-22	Sequence 22, Appli
43	32.4	1.7	713	3	US-08-532-896-26	Sequence 26, Appli
44	32	1.7	2570	2	US-09-056-075-2	Sequence 2, Appli
45	32	1.7	3291	1	US-08-021-601-1	Sequence 1, Appli

RESULT 1

US-08-232-463-14

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 1800 Diagonal Road, Suite 500

; CITY: Alexandria

; STATE: VA

; COUNTRY: USA

; ZIP: 22313-0299

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/232,463

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/935,313

; FILING DATE:

; APPLICATION NUMBER: EP 91 114 300.6

; FILING DATE: 26-AUG-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 30472/114 IMMU

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)836-9300

; TELEFAX: (703)683-4109

; TELEX: 899149

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 7218 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: pTZgpt-F1s

US-08-232-463-14

[illegible]

2
RESULT
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match 3.1%; Score 57.4; DB 1; Length 7218;
Best Local Similarity 4.3%; Pred. No. 6.7e-07;
Matches 16; Conservative 211; Mismatches 142; Indels 0; Gaps

Qy 257 agaaagcacaagaagcacagacagatccctgggagaaatgccgcgcgcattcttggt 316
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1407 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1348
Qy 317 catgatgcttcctaaagatggcgagtgctgctgctgctcccgctgtgagggagacattagaaaatga 376
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1347 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1288
Qy 377 atgatgtgtctctaaagatggcgagtgagcagaaacagatccctgtgtggtattatttga 436
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1287 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1228
Qy 437 acgggattacagattgaaatgaatcaacaaagtgcattaccaatgagaggaacag 496
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1227 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1168
Qy 497 acgagaaaatcttgatggcttcacaagacatgcacaacaaaatggaaactgtgatga 556
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1167 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1108
Qy 557 catgaggcagccaagctggyggaggagataaccacgggcagagggctcaggattctgccc 616
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1107 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1048
Qy 617 tgctgccta 625
      ||| ||| |
Db 1047 TGCAGCCAA 1039

RESULT 3
US-08-680-327-2/c
; Sequence 2, Application US/08680327
; Patent No. 5859321
; GENERAL INFORMATION:
; APPLICANT: Staskawicz, Brian S., Oldroyd, Giles Edward,
; APPLICANT: Salmeron, John M., Rommens, Calus
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT
; TITLE OF INVENTION: PATHOGEN RESISTANCE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Klarquist Sparkman Campbell Leigh &
; ADDRESSEE: Whinston
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,327
; FILING DATE: July 11, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:

```



```

; FILING DATE: April 13, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/254001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 157:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-310-912A-157

Query Match 2.4%; Score 45.4; DB 2; Length 5134;
Best Local Similarity 49.7%; Pred. No. 0.0021;
Matches 153; Conservative 4; Mismatches 148; Indels 3; Gaps 2;

QY 1303 coatnysacctaaacaaatctaaacttgtaattccttgaaatgtcaggnccatcacatrrt 1362
||||: || || || || || || || || || || || || || || || || || || || || ||
Db 806 CCATTGTCTTAGGGAAGCTATAAATCTGAAGAACTCTGGTAC-TTTCAGACAAAATAATCTT 748

QY 1363 tctcttgctgagaagctcttctcttcttaantctagaatgatgaaagttttgaa 1422
||||: || || || || || || || || || || || || || || || || || || || || ||
Db 747 TACCTCTGCCTCAACAACTTAAGCAATTCCTTGACATCAGAGAGATAGTGTGTAATA 688

QY 1423 taagttgactattcttctcatgcaagaaggacacatgatgagattcatcacatga 1482
||||: || || || || || || || || || || || || || || || || || || || || ||
Db 687 CCAAGCAGGAATATACCCCTTGACAGAGAAACCAAGTGTAGTTTATCTTGACACAG 628

QY 1483 gacagcaataactaaagtgttaattgattataagagtttagataataatgaaatgca 1542
||||: || || || || || || || || || || || || || || || || || || || || ||
Db 627 TTCAGCGAGATCTGTAAGTCTATATATGCATATTCGCGTACTGTACAAATATCTACAAGTAA 568

QY 1543 agakccacagaggaatgtttatggggcagctttgtaagcctgggagtgaaagaaagc 1602
||||: || || || || || || || || || || || || || || || || || || || || ||
Db 567 AGAGCCAAATGTGCAATATTTTGTCTTAACACTG--AGATTGGTTAAGAGATAAGAC 510

QY 1603 aggggaacc 1610
| |||||
Db 509 ACTGAACC 502

RESULT 6
PCT-US95-04589-157/c
; Sequence 157, Application PC/TUS9504589
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE AND USES THEREOF
; NUMBER OF SEQUENCES: 201
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04589
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,360
; FILING DATE: 13-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 157:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
PCT-US95-04589-157

Query Match 2.4%; Score 45.4; DB 5; Length 5134;
Best Local Similarity 49.7%; Pred. No. 0.0021;
Matches 153; Conservative 4; Mismatches 148; Indels 3; Gaps 2;

QY 1303 coatnysacctaaacaaatctaaacttgtaattccttgaaatgtcaggnccatcacatrrt 1362
||||: || || || || || || || || || || || || || || || || || || || || ||
Db 806 CCATTGTCTTAGGGAAGCTATAAATCTGAAGAACTCTGGTAC-TTTCAGACAAAATAATCTT 748

QY 1363 tctcttgctgagaagctcttctcttcttaantctagaatgatgaaagttttgaa 1422
||||: || || || || || || || || || || || || || || || || || || || || ||
Db 747 TACCTCTGCCTCAACAACTTAAGCAATTCCTTGACATCAGAGAGATAGTGTGTAATA 688

QY 1423 taagttgactattcttctcatgcaagaaggacacatgatgagattcatcacatga 1482
||||: || || || || || || || || || || || || || || || || || || || || ||
Db 687 CCAAGCAGGAATATACCCCTTGACAGAGAAACCAAGTGTAGTTTATCTTGACACAG 628

QY 1483 gacagcaataactaaagtgttaattgattataagagtttagataataatgaaatgca 1542
||||: || || || || || || || || || || || || || || || || || || || || ||
Db 627 TTCAGCGAGATCTGTAAGTCTATATATGCATATTCGCGTACTGTACAAATATCTACAAGTAA 568

QY 1543 agakccacagaggaatgtttatggggcagctttgtaagcctgggagtgaaagaaagc 1602
||||: || || || || || || || || || || || || || || || || || || || || ||
Db 567 AGAGCCAAATGTGCAATATTTTGTCTTAACACTG--AGATTGGTTAAGAGATAAGAC 510

QY 1603 aggggaacc 1610
| |||||
Db 509 ACTGAACC 502

RESULT 7
US-08-327-1/c
; Sequence 1, Application US/08680327
; Patent No. 5859321
; GENERAL INFORMATION:
; APPLICANT: Staskawicz, Brian S., Oldroyd, Giles Edward,
; APPLICANT: Salmeron, John M., Rommens, Caius
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT
; TITLE OF INVENTION: PATHOGEN RESISTANCE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist Sparkman Campbell Leigh &
; ADDRESS: Whinston
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
```



```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/586,509
; FILING DATE: 11-JAN-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Perkins, Patricia Anne
; REGISTRATION NUMBER: 34,693
; REFERENCE/DOCKET NUMBER: 2841
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14507 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: No. 6027915 Relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Chinese hamster
; IMMEDIATE SOURCE:
; CLONE: 2A5-3 lambda CHO sequence
US-08-785-150-1

Query Match 1.9%; Score 35.8; DB 3; Length 14507;
Best Local Similarity 46.9%; Pred. No. 2.9;
Matches 112; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 870 gtgttctgcccaggatctgtgaacagctgaggagcattctcaagatctttccagggt 929
Db 10959 GGGTCTTCCACAGTGGATGATGCTCATCAAGGGAATTTATTCACCCCTGGTGCTAGTG 10900

QY 930 tatactactagcacacagcatgatcattacagagtgagtgaaattatcttaatacaacatcatcct 989
Db 10899 GCTACAATAGCCGGCTAGAGGCGACAGACAGCCTTCTTAATCAGGTCCTGTGTGTAC 10840

QY 990 cagtgtcttgccatactgaattcatcttcccaactttgtgcccatttccaagaacctca 1049
Db 10839 GAGCATGACTAAGCGCCAGTATACAGCGCAACCTGTTTGAGAAATTTAAAGTGTCTCA 10780

QY 1050 aaatgtcattccatttaataccagagattaaacttttttttaacctgggaagaattcaa 1108
Db 10779 CACAGTAGTTCAGTAGTAGACAGACAGAAACATTTATTTTATTTGGAAAAATATAA 10721

RESULT 10
US-08-370-319C-12/c
; Sequence 12, Application US/08370319C
; Patent No. 5856091
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; APPLICANT: Traversari, Catia; W lfel, Thomas; Coullie, Pierre;
; APPLICANT: Boon-Falleur, Thierry; De plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE TU
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,319C
; FILING DATE: 10-JANUARY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/272,351
; FILING DATE: 8-JULY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/032,978
; FILING DATE: 18-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5856091man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5377.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; OTHER INFORMATION: The sequence is preceded by an
; OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3
; OTHER INFORMATION: kilobases
US-08-370-319C-12

Query Match 1.9%; Score 35; DB 2; Length 4129;
Best Local Similarity 48.7%; Pred. No. 2.4;
Matches 95; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 1025 ttttgcgccattctcaagacctcaaaatgtcattccattatatacacaggattacttt 1084
Db 2703 TCTTTTCTACATACTAATAATTTTGAATATGATATGATATATATATATATATATATAT 2644

QY 1085 ttttttaacctgggaattcaattgttacatgcagctatgggaatttaattacatatatt 1144
Db 2643 AACTCTGTGTTTAAAGTGTTCATAGCACCATGACTAGTGGCTAAGCCATCAGACAG 2584

QY 1145 tgttttccagtgcaagatgactaaagtccttccctccctcccttggttgttttttttc 1204
Db 2583 CACAGCTCTATACAAATAAATAGTACTTCTTCCCTCCCTCTCTATATATATCTGCATC 2524

QY 1205 cagtataaaggtaaa 1219
Db 2523 CATTAAGGCAAA 2509

RESULT 11
US-09-224-834-12/c
; Sequence 12, Application US/09224834
; Patent No. 6201111
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; APPLICANT: Traversari, Catia; W lfel, Thomas; Coullie, Pierre;
; APPLICANT: Boon-Falleur, Thierry; De plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; SOFTWARE: IBM PS/2

```

```

: GENERAL INFORMATION:
:
: APPLICANT: Benton, Bret
:
: APPLICANT: Lee, Ving J.
:
: APPLICANT: Malouin, Francois
:
: APPLICANT: Martin, Patrick K.
:
: APPLICANT: Schmid, Molly B.
:
: APPLICANT: Sun, Dongxu
:
: TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
:
: TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
:
: TITLE OF INVENTION: TARGET GENES
:
: NUMBER OF SEQUENCES: 111
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Lyon & Lyon
:
: STREET: 633 West Fifth Street
:

```


Search completed: July 31, 2001, 07:20:07
Job time: 17569 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:25:26 ; Search time 478.87 Seconds
(without alignments)
2454.595 Million cell updates/sec

Title: US-09-402-713A-3
Perfect score: 1872
Sequence: 1 agaagctgcatcagaaaaa.....caataagaattacaaga 1872

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues 1460202
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1867	99.7	1872	19	AAV62428
2	1772.4	94.7	2037	19	AAV62427
3	1772.4	94.7	3582	19	AAV62430
4	1739.2	92.9	3112	21	AAA06687
c 5	1734	92.6	2426	21	AAA06689
c 6	1715	91.6	2229	21	AAA06688
7	820	43.8	820	19	AAV62429
c 8	812	43.4	812	21	AAA06690
9	524	28.0	597	20	AAV62486
10	513.2	27.4	718	21	AAA06545
11	414.2	22.1	437	21	AAC06768

c 12	330	17.6	359	20	AA233445	Human prostate can
c 13	288.4	15.4	301	21	AA06520	Human immunogenic
c 14	136	7.3	936	22	AAF58252	Oligonucleotide D1
c 15	136	7.3	936	22	AAF58254	Oligonucleotide D1
c 16	136	7.3	936	22	AAF58257	Oligonucleotide D1
c 17	136	7.3	936	22	AAF58259	Oligonucleotide D2
c 18	136	7.3	936	22	AAF58262	Oligonucleotide D2
c 19	136	7.3	936	22	AAF58255	Oligonucleotide D1
c 20	134.4	7.2	936	22	AAF58252	Oligonucleotide D1
c 21	134.4	7.2	936	22	AAF58254	Oligonucleotide D1
c 22	134.4	7.2	936	22	AAF58257	Oligonucleotide D1
c 23	134.4	7.2	936	22	AAF58259	Oligonucleotide D2
c 24	134.4	7.2	936	22	AAF58262	Oligonucleotide D2
c 25	134.4	7.2	936	22	AAF58255	Oligonucleotide D1
c 26	100	5.3	123	19	AAV33791	Prostate cancer an
c 27	50.8	2.7	244	22	AAF58238	Oligonucleotide D1
c 28	47.6	2.5	244	22	AAF58238	Oligonucleotide D1
c 29	47	2.5	10968	19	AAV17789	Tomato Prf genomic
c 30	45.4	2.4	5475	19	AAV17777	Tomato Prf cDNA.
c 31	40	2.1	796	20	AAV37484	Human secreted pro
c 32	39	2.1	2206	21	AA12390	Human interferon-b
c 33	39	2.1	10325	20	AAV76516	Plasmid containing
c 34	37.6	2.0	1297	13	AAQ21645	3' coding sequence
c 35	37.6	2.0	1770	14	AAQ37684	P.falciiparum antig
c 36	37.6	2.0	1770	14	AAQ37124	P.falciiparum antig
c 37	37.4	2.0	1255	22	AAF72762	Human prostate can
c 38	37.4	2.0	1325	21	AAF15730	Human prostate can
c 39	37.4	2.0	1325	22	AAF72787	Human prostate can
c 40	37.2	2.0	1751	21	AAV62428	Human secreted pro
c 41	36.8	2.0	5804	18	AAV74560	Staphylococcus aur
c 42	36.4	1.9	2643	15	AAQ66096	S. cerevisiae Plc1
c 43	36.2	1.9	1451	21	AAF21739	Human breast and o
c 44	36.2	1.9	1517	21	AAC34792	Arabidopsis thalia
c 45	36.2	1.9	1630	20	AAV13310	Enterococcus faeca

ALIGNMENTS

RESULT 1	
ID AAV62428	standard; cDNA; 1872 BP.
XX	
AC AAV62428;	
XX	
DT 30-DEC-1998	(first entry)
XX	
DE	Prostate cancer antigen (PCA3) cDNA splice variant 2.
XX	
KW	Prostate cancer antigen cDNA splice variant 2; PCA3; prostatic cancer;
KW	PC; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO9845420-A1.
XX	
PD	15-OCT-1998.
XX	
PF	09-APR-1998; 98WO-CA00345.
XX	
PR	10-APR-1997; 97US-0041836.
XX	
PA	(DIAG-) DIAGNOCURE INC.
XX	
PI	Bussemakers MJG;
XX	
DR	WPI; 1998-568347/48.
XX	
PT	New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT	prevention and treatment of prostatic cancer
XX	
PS	Claim 4; Pages 76-77; 11lpp; English.
XX	

CC The present sequence represents the prostate cancer antigen (PCa3)
CC cDNA splice variant 2 sequence comprising of exons 1, 3, 4a and
CC 4b of the PCa3 gene. The PCa3 cDNA splice variant 2 sequence,
CC isolated from a human primary prostatic tumour tissue cDNA library,
CC was found in approximately 65% of the cDNA clones isolated. The
CC invention claims for PCa3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCa3 protein. The
CC antibodies are claimed to be useful for detecting PCa3 protein in
CC immunoassay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCa3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCa3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCa3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCa3 expression allows correlation with the grade of
CC tumour. PCa3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (antagonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 1872 BP; 567 A; 389 C; 369 G; 539 T; 8 other;

Query Match 99.7%; Score 1867; DB 19; Length 1872;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1872; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agaagctggcatcagaaaaacagagggagatttggctgcagccgagggagaccag 60
DB 1 agaagctggcatcagaaaaacagagggagatttggctgcagccgagggagaccag 60
QY 61 gaagatctgcattgtgggaagacacctgatgacagaggtgagaaataagaagacctgt 120
DB 61 gaagatctgcattgtgggaagacacctgatgacagaggtgagaaataagaagacctgt 120
QY 121 gactttaccatctgagggccacacatctgctgaaatggagataattaacatcactagaac 180
DB 121 gactttaccatctgagggccacacatctgctgaaatggagataattaacatcactagaac 180
QY 181 agcaagatgacataataatgtctaaagttagtgcattgttttgcatttccagcccttt 240
DB 181 agcaagatgacataataatgtctaaagttagtgcattgttttgcatttccagcccttt 240
QY 241 aaatattccacacacaggaagacaaaaaggaagcacagagatccctggagaaatgcc 300
DB 241 aaatattccacacacaggaagacaaaaaggaagcacagagatccctggagaaatgcc 300
QY 301 ggcgcacattgggtcatcgatgagcctgcgccctgtgctggctcccgcttggaggaa 360
DB 301 ggcgcacattgggtcatcgatgagcctgcgccctgtgctggctcccgcttggaggaa 360
QY 361 ggacattgaaatgaattgtgttccctaaaggatggcgaggagaaacagatcctgtt 420
DB 361 ggacattgaaatgaattgtgttccctaaaggatggcgaggagaaacagatcctgtt 420
QY 421 gtggattatttgaacgggattacagatttgaatgaagtcaaaaagtgagcattacc 480
DB 421 gtggattatttgaacgggattacagatttgaatgaagtcaaaaagtgagcattacc 480
QY 481 aatgagagaaacacagacagaaaaatttgtatgcttcaagaagacatgcaaaaacaaa 540
DB 481 aatgagagaaacacagacagaaaaatttgtatgcttcaagaagacatgcaaaaacaaa 540
QY 541 tggaaactgtgtgacatgagggcgaagctggggaggagataaccacggggcagagg 600
DB 541 tggaaactgtgtgacatgagggcgaagctggggaggagataaccacggggcagagg 600
QY 601 gtcaagatttgcctgtcgtcctaactgtcgttccataaaccataatttcattttc 660
DB 601 gtcaagatttgcctgtcgtcctaactgtcgttccataaaccataatttcattttc 660
QY 661 taacctcaaaaacagctgttgtaattctgtatctctacggttctctctggggccaaca 720

DB 661 taacctcaaaaacagagctgttgtaattctgtatctacggttctctctggggccaaca 720
QY 721 ttctccatatccacgcaacacactcattttaattatttagttccagatctgtactgtac 780
DB 721 ttctccatatccacgcaacacactcattttaattatttagttccagatctgtactgtac 780
QY 781 cttctcacactgtagaataacattactcttttgcataagaccctctgtgtgtgcct 840
DB 781 cttctcacactgtagaataacattactcttttgcataagaccctctgtgtgtgcct 840
QY 841 aatatgtactgtactgttttttcttaaggagttgttctggccagggtgactgtgaacaggc 900
DB 841 aatatgtactgtactgttttttcttaaggagttgttctggccagggtgactgtgaacaggc 900
QY 901 tgggaagcactcagaagatttttccaggggttatacttactagcacacagatgatattac 960
DB 901 tgggaagcactcagaagatttttccaggggttatacttactagcacacagatgatattac 960
QY 961 ggagtgaaattatcctaatacacaatcctcagtgcttctggccatactgaaattcatttc 1020
DB 961 ggagtgaaattatcctaatacacaatcctcagtgcttctggccatactgaaattcatttc 1020
QY 1021 ccactttgtgccattctcaagacacctcaaaatgtctatccatttaataatcaccaggattaa 1080
DB 1021 ccactttgtgccattctcaagacacctcaaaatgtctatccatttaataatcaccaggattaa 1080
QY 1081 ctttttttttaacctggaagaattcaattgtttacatgcagctatgggaatttaattacat 1140
DB 1081 ctttttttttaacctggaagaattcaattgtttacatgcagctatgggaatttaattacat 1140
QY 1141 atttgttttccagtgcaagatgactaagctctttatccctccctttgtttgatttt 1200
DB 1141 atttgttttccagtgcaagatgactaagctctttatccctccctttgtttgatttt 1200
QY 1201 ttccagtgataaagttaaaatgcttagcctgttactgaggtgtgtatcacagcacgctct 1260
DB 1201 ttccagtgataaagttaaaatgcttagcctgttactgaggtgtgtatcacagcacgctct 1260
QY 1261 cccatccctccagcttactgtctacatccatccatccctccatccctccatccatccatcc 1320
DB 1261 cccatccctccagcttactgtctacatccatccatccctccatccctccatccatccatcc 1320
QY 1321 atctaacttgtaattccttggaacatgctcaggncatcacatrttctcttgcctgagaagc 1380
DB 1321 atctaacttgtaattccttggaacatgctcaggncatcacatrttctcttgcctgagaagc 1380
QY 1381 tcttctgtctcttaantctagaatgataaaagttttgaataagttgactatcttact 1440
DB 1381 tcttctgtctcttaantctagaatgataaaagttttgaataagttgactatcttact 1440
QY 1441 tcatcaagaagaaggacacacatgagattcactcactcagatgagacacaaactactaaag 1500
DB 1441 tcatcaagaagaaggacacacatgagattcactcactcagatgagacacaaactactaaag 1500
QY 1501 tgtaatttgattataagagtttagataaaatatgaaatgcaagakccacagaggaatg 1560
DB 1501 tgtaatttgattataagagtttagataaaatatgaaatgcaagakccacagaggaatg 1560
QY 1561 tttaaggggcagcttggtaagcctgggagtgtgaamaaggcagggaacctcatagtc 1620
DB 1561 tttaaggggcagcttggtaagcctgggagtgtgaamaaggcagggaacctcatagtc 1620
QY 1621 ttataataatacttcttct 1680
DB 1621 ttataataatacttcttct 1680
QY 1681 ttcagcagtgcaaatcccccaaggtaacctttatccatttccatggtagtcgcttag 1740
DB 1681 ttcagcagtgcaaatcccccaaggtaacctttatccatttccatggtagtcgcttag 1740
QY 1741 aatttggcaaatcactactgttgcacttctcacttctcacttctcacttctcacttctcact 1800


```
Db 1219 gtcattccatataatcacaggattaaacttttttttttaacctggaagaattcaaatgtta 1278
QY 1114 catgagcatatgggaatttaattacataattttgttttccagtgcaaaagatgactaagtcc 1173
Db 1279 catgagcatatgggaatttaattacataattttgttttccagtgcaaaagatgactaagtcc 1338
QY 1174 ttatccctccctctgtgtgtgttttttttccagtataaagttaaactcttagccttgt 1233
Db 1339 ttatccctccctctgtgtgtgttttttttccagtataaagttaaactcttagccttgt 1398
QY 1234 actgaggtgtatcacagcagcctctccctccctccctccctccctccctccctccctccct 1293
Db 1399 actgaggtgtatcacagcagcctctccctccctccctccctccctccctccctccctccct 1458
QY 1294 caaccctccctccctccctccctccctccctccctccctccctccctccctccctccctcc 1353
Db 1459 caaccctccctccctccctccctccctccctccctccctccctccctccctccctccctcc 1518
QY 1354 atacatrrtctctctgctgagagctctctctgtctctctctctctctctctctctctctct 1413
Db 1519 atacatrrtctctgctgagagctctctctgtctctctctctctctctctctctctctct 1578
QY 1414 agtttgaataagttgactatcttactctcatgcaagaaggagacacatatgagattcatc 1473
Db 1579 agtttgaataagttgactatcttactctcatgcaagaaggagacacatatgagattcatc 1638
QY 1474 atcacatgagacagcaaaactactaaaagttaatttgattatagaagtttagataaata 1533
Db 1639 atcacatgagacagcaaaactactaaaagttaatttgattatagaagtttagataaata 1698
QY 1534 tgaattgcaagkccacagaggaattgtttatgggcacgttttgaagcctgggatgta 1593
Db 1699 tgaattgcaagkccacagaggaattgtttatgggcacgttttgaagcctgggatgta 1758
QY 1594 agmaaggcagggaacctcatagttctctataataataataataataataataataataata 1653
Db 1759 agmaaggcagggaacctcatagttctctataataataataataataataataataataata 1818
QY 1654 acaatatcaacaagcttttccagaaattcatgcagtgcaaatcccccaagtaaccttt 1713
Db 1819 acaatatcaacaagcttttccagaaattcatgcagtgcaaatcccccaagtaaccttt 1878
QY 1714 atccatttcattgtagtgctgttagaattttggcaaatcactactggtcacttatctca 1773
Db 1879 atccatttcattgtagtgctgttagaattttggcaaatcactactggtcacttatctca 1938
QY 1774 acttgagatggtttgtctctgttagtttaattgaaagaaataggcgcaactcttgtgagcca 1833
Db 1939 acttgagatggtttgtctctgttagtttaattgaaagaaataggcgcaactcttgtgagcca 1998
QY 1834 ctttaggggttactctctggcaataaagaatttacaaaga 1872
Db 1999 ctttaggggttactctctggcaataaagaatttacaaaga 2037
```

RESULT 3

AAV62430
ID AAV62430 standard; cDNA; 3582 BP.

XX AAV62430;

XX AAV62430;

DT 30-DEC-1998 (first entry)

DE Prostate cancer antigen (PCA3) wild-type cDNA.

XX Prostate cancer antigen cDNA; PCA3; prostatic cancer;

KW PC; ds.

XX Homo sapiens.

OS Prostate cancer antigen cDNA; PCA3; prostatic cancer;

XX PC; ds.

```
FT FT /*tag= a  
FT FT /product= "PCA3 protein"  
FT FT 983..987  
FT FT /*tag= b  
FT FT polyA_signal  
FT FT 2041..2046  
FT FT /*tag= c  
FT FT polyA_signal  
FT FT 2597..2602  
FT FT /*tag= d  
FT FT polyA_signal  
FT FT 3494..3496  
FT FT /*tag= e  
XX  
PN WO9845420-A1.  
XX  
XX 15-OCT-1998.  
XX  
XX 09-APR-1998; 98WO-CA00346.  
XX  
XX 10-APR-1997; 97US-0041836.  
XX  
XX (DIAG-) DIAGNOCURE INC.  
XX  
XX Bussemakers MJG;  
XX  
XX WPI; 1998-568347/48.  
XX  
XX P-PSDB; AAW79738.  
XX  
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,  
XX prevention and treatment of prostatic cancer  
XX  
XX Claim 3; Fig 5B-5F; lllpp; English.  
XX  
XX The present sequence represents the prostate cancer antigen (PCA3)  
XX wild-type cDNA sequence comprising of exons 1, 2, 3, 4a-4d of the  
XX PCA3 gene. The invention claims for PCA3 cDNA variants and the  
XX proteins they encode. The invention also claims for antibodies  
XX against PCA3 protein. The antibodies are claimed to be useful for  
XX detecting PCA3 protein in immunoassay tests, for diagnosing, assessing  
XX and prognosing of prostatic cancer (PC). Antibodies, optionally  
XX coupled to a cytotoxin or radioisotope, and nucleic acids antisense  
XX to PCA3 cDNA are claimed to be useful for treating PC, while determining  
XX elevated levels of PCA3 (as RNA or protein) is useful for detecting a  
XX predisposition to development of PC, e.g. in prenatal tests. Detecting  
XX PCA3 protein allows differentiation between malignant and benign  
XX prostatic disease, and the level of PCA3 expression allows correlation  
XX with the grade of tumour. PCA3 protein and its fragments are also  
XX claimed to be useful in vaccines for preventing PC; in drug screens  
XX for identifying specific (antagonists (potentially useful  
XX therapeutically) and for studying protein-DNA interactions.  
XX  
XX Sequence 3582 BP; 1052 A; 788 C; 679 G; 1063 T; 0 other;
```

Query Match 94.7%; Score 1772.4; DB 19; Length 3582;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1770; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```
QY 94 cagaggtgagaataagaaggctgctgactttaccatctgagggccacacatctgtctgaa 153  
Db 281 cagggtgagaaataagaaggctgctgactttaccatctgagggccacacatctgtctgaa 340  
QY 154 atggagataattacatcactgaaacagagatgacaataatgtctaagttagtac 213  
Db 341 atggagataattacatcactgaaacagagatgacaataatgtctaagttagtac 400  
QY 214 atgtttttgcacatttccagcccttttaataatccacacagagaaacacaaaggaa 273  
Db 401 atgtttttgcacatttccagcccttttaataatccacacagagaaacacaaaggaa 460  
QY 274 gcacagagatccctggagaaatgccggccgcctcttggttcacatgcgatgagcctcgcc 333  
Db 461 gcacagagatccctggagaaatgccggccgcctcttggttcacatgcgatgagcctcgcc 520  
QY 334 ctgtgcctggtcccgcttgtgaggggaagacattagaaaaatgaattgattgtgttccttaa 393
```

Db 521 ctgtgcttgcctgtgcgctgtgagggaaggacattagaaaaatgattgattgcttcttaa 580
Qy 394 aggatggcaggaataaacagatcctgttctgtgatatatttttgaacggattacagatttg 453
Db 581 agga tggcaggaataaacaga cctgttctgtgata tttatttgaacggattacagatttg 640
Qy 454 aaatgaagtccaaagttagcattaccatgagaggaataaacagacagagaaatcttgatg 513
Db 641 aaatgaagtccaaagttagcattaccatgagaggaataaacagacagagaaatcttgatg 700
Qy 514 gcttcacaagacatgcacaacaaaa tggaaatctgtgatactgtgacatgaggcgcaagct 573
Db 701 gcttcacaagacatgcacaacaaaa tggaaatctgtgatactgtgacatgaggcgcaagct 760
Qy 574 gggagagagataaacacagggcgagaggtcagattcttgcgcctgtcctcaaacgtgac 633
Db 761 ggggaggagataaacacagggcgagaggtcagattcttgcgcctgtcctcaaacgtgac 820
Qy 634 gtctataaccataatcttcatattcttaacctcaaaccaaaagctgtgtaatacttga 693
Db 821 gtctataaccataatcttcatattcttaacctcaaaccaaaagctgtgtaatacttga 880
Qy 694 tctctacaggttctctctgggcccacattctccatatatccagccacactcatctttaat 753
Db 881 tctctacaggttctctctgggcccacattctccatatatccagccacactcatctttaat 940
Qy 754 atttagttccagatctgtactgtgaccttctctacactgtagaataacattactcatctt 813
Db 941 atttagttccagatctgtactgtgaccttctctacactgtagaataacattactcatctt 1000
Qy 814 gtccaaagaccctctgtgtgctgctctaata tggatgagctgttttctcctaaggagtgt 873
Db 1001 gtccaaagaccctctgtgtgctgctctaata tggatgagctgttttctcctaaggagtgt 1060
Qy 874 tctggcccagggtatctgtgaacaggctgggaagcatctcaagatcttccagggttata 933
Db 1061 tctggcccagggtatctgtgaacaggctgggaagcatctcaagatcttccagggttata 1120
Qy 934 cttaactagcacacagcatgatactacagagtgagtgatcttaataacatacatctcag 993
Db 1121 cttaactagcacacagcatgatactacagagtgagtgatcttaataacatacatctcag 1180
Qy 994 gtcttgcccatactgaaattcatctccacttccactttgtgcccattctcaagacctcaaat 1053
Db 1181 gtcttgcccatactgaaattcatctccacttccactttgtgcccattctcaagacctcaaat 1240
Qy 1054 gtcatctcatttaataatcacaggattaaacttttttttaacctggagaataattcaatgtta 1113
Db 1241 gtcatctcatttaataatcacaggattaaacttttttttaacctggagaataattcaatgtta 1300
Qy 1114 catgagctatgggaatttaattacataattttgttttccagtgcaaatgactaagctcc 1173
Db 1301 catgagctatgggaatttaattacataattttgttttccagtgcaaatgactaagctcc 1360
Qy 1174 tttaacctcccccttgttgaatttttttccagataaagttaaagtcttagccttgt 1233
Db 1361 tttaacctcccccttgttgaatttttttccagataaagttaaagtcttagccttgt 1420
Qy 1234 actgaggtgtatacagacagcctctcccccattccctccagccttatctgtcatcaccat 1293
Db 1421 actgaggtgtatacagacagcctctcccccattccctccagccttatctgtcatcaccat 1480
Qy 1294 caacccctcccatnysactaaacaaaaataacttgaattcttgaactgacatgcaggnc 1353
Db 1481 caacccctcccatcaccataacaaaaataacttgaattcttgaactgacatgcaggac 1540
Qy 1354 atacattttctctgtcgtgagagctctctctgtcttcttaantctagaatgatgtaa 1413
Db 1541 atacattttctctgtcgtgagagctctctctgtcttcttaantctagaatgatgtaa 1600
Qy 1414 agttttgaataagttgactatcttacttcctgcaagaagggacacatatgagattcatc 1473

Db 1601 agttttgaataagttgactatcttacttcctgcaagaaggacacatatgagattcatc 1660
Qy 1474 atccatcagacagacaaaataactaaagtgttaatttgattataaagatttagataaata 1533
Db 1661 atccatcagacagacaaaataactaaagtgttaatttgattataaagatttagataaata 1720
Qy 1534 tgaataatgcaagakccacagagaggaatttttatggggcacgctttgtaagcctgggatgta 1593
Db 1721 tgaataatgcaagagccacagaggaatttttatggggcacgctttgtaagcctgggatgta 1780
Qy 1594 agmaagcaggaggaacctcatagtcttataataataacttctctatctctatc 1653
Db 1781 agcaagcagggaacctcatagtcttataataataacttctctatctctatc 1840
Qy 1654 acaatatccaaacagcttttccacagaattcatgcagtcagtcacaaatccccaaagtaaccttt 1713
Db 1841 acaatatccaaacagcttttccacagaattcatgcagtcagtcacaaatccccaaagtaaccttt 1900
Qy 1714 atccatttcattgtgagtcgcttttagaatttttggcaaatcacttggtcaacttatctca 1773
Db 1901 atccatttcattgtgagtcgcttttagaatttttggcaaatcacttggtcaacttatctca 1960
Qy 1774 accttgagatgtgttctgtctctgttagttaaattgaaagaataaggcactctctgtagcaca 1833
Db 1961 accttgagatgtgttctgtctctgttagttaaattgaaagaataaggcactctctgtagcaca 2020
Qy 1834 cttttaggttcaactcctctgcaataaagaatttacaaga 1872
Db 2021 cttttaggttcaactcctctgcaataaagaatttacaaga 2059
RESULT 4
AAAA06687
ID AAA06687 standard; cDNA; 3112 BP.
XX
AC AAA06687;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:468.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
(CORI-) CORIXA CORP.
PI Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 259-260; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides

QY 1444 tgcaagaaggagacacatatgagattcatcatcacatgagacagcaaaataactataaagtgt 1503
 |||||
 Db 424 TGCAAGAAGGACACATATGAGATTTCATCATCATGACACAGCAAAATACTAAAGAGTGT 365
 QY 1504 aatttgattataagagcttagataataatatataatgaatgcaagakccacagaggaatgtttt 1563
 |||||
 Db 364 AATTTGATTATAGAGCTTTAGATATAATATATAATGAAATGCAAGAGCCACAGAGGAATGTTT 305
 QY 1564 atggggcagctgttgtaagctggagtgtaagaaagcagaggaacccatcatagtattcta 1623
 |||||
 Db 304 ATGGGGCAGCTTTGTGAAGCTTGGGATGTGAAGCAAGGAGGAGAACTCATAGTATCTTA 245
 QY 1624 tataataacttcatct 1683
 |||||
 Db 244 TATAATATACTTCATTTCT 185
 QY 1684 atgcagtcgaataatccccaaagtaaacctttatccattcatcattcatcattcatcatt 1743
 |||||
 Db 184 ATGCAGTGCAAATCCCAAGGTAACCTTTATCCATTTTCATGAGTGGCTTTAGAAT 125
 QY 1744 ttggcaaatcatctgctcacttctctcaactttgagatgtgtttgtctctgttaattaa 1803
 |||||
 Db 124 TTGGCAATATCATCTGGTCACTTCTCACTTTTGAGATGTGTTGTCTTGTAGTTAA 65
 QY 1804 ttgaagaataaggcgaactctgtgagccacttttaggttcaactcctcctggcaataaagaat 1863
 |||||
 Db 64 TTGAAGAATAAGGGCACCTTTGTGAGCCACTTTAGGGTTCACTCTCGGCATAAAGANT 5
 QY 1864 ttac 1867
 |||||
 Db 4 TTAC 1.

RESULT 6

AA06688/c
 ID AA06688 standard; cDNA; 2229 BP.

XX
 AC AAA06688;
 DT
 TT 13-JUN-2000 (first entry)
 XX
 DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:469.
 XX
 KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
 KW immunogenic; cytostatic; vaccine; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200004149-A2.

XX
 PD 27-JAN-2000.
 XX
 PF 14-JUL-1999; 99WO-US15838.

XX
 PR 14-JUL-1999; 98US-0115453.
 PR 14-JUL-1998; 98US-0116134.
 PR 23-SEP-1998; 98US-0159812.
 PR 23-SEP-1998; 98US-0159822.
 PR 15-JAN-1999; 99US-0232149.
 PR 15-JAN-1999; 99US-0232880.
 PR 09-APR-1999; 99US-0288946.

XX
 PA (CORI-) CORIXA CORP.
 XX

XX
 PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
 XX

XX
 DR WPI; 2000-171268/15.
 XX

XX
 PT New polypeptide useful for treating and diagnosing prostate cancer
 PT comprises an immunogenic portion of prostate tumor protein -
 XX

XX
 PS Claim 1; Page 260-261; 263pp; English.

XX

CC The present invention describes isolated polypeptides, comprising an
 CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
 CC and polynucleotides encoding them have cytostatic activity and can be
 CC used in vaccines and in gene therapy. The polypeptides and
 CC polynucleotides encoding them, antigen presenting cells which express
 CC the polypeptides, antibodies against the polypeptides and vaccines
 CC comprising them can be used for inhibiting the development of prostate
 CC cancer in a patient. The polypeptides can be used to generate antibodies
 CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
 CC the polynucleotides encoding the polypeptides can be used as a probe or
 CC to modulate the expression of the polypeptides. AA06241 to AA06691 and
 CC AA082000 to AA082020 represent sequences used in the exemplification of
 CC the present invention.

XX
 SQ Sequence 2229 BP; 654 A; 447 C; 481 G; 647 T; 0 other;

Query Match 91.6%; Score 1715; DB 21; Length 2229;

Best Local Similarity 98.3%; Pred. No. 0;

Matches 1779; Conservative 4; Mismatches 21; Indels 5; Gaps 5;

QY 65 atctgcattggtgggaagacctgatgatacacagaggtgagaaataagaaaggctgtgact 124

Db 1808 ATTTGTCTCTCTCAAAATGGTCTGATGTATTTCCAGGTGAGAAATAAGAAAGGCTGTGACT 1749

QY 125 ttaccatctgagccacacacatctgctgaatgagataattaacatcactagaacagca 184

Db 1748 TTACCATCTGAGGCCACACATCTGCTGAATGGAGATAATTACATCATTAGAACAGCA 1689

QY 185 agatgacaataatagtctaaagttagtgacatgtttttgcacattttccagcccttttaaa 244

Db 1688 AGATGACAATAATATCTCTAAGTAGTAGACATGTTTTGACATTTCCAGCCCTTTAAT 1629

QY 245 atccacacacaggaagcacaaggaagcagacagatccctgggagaaatgcgcgcgc 304

Db 1628 ATCCACACACAGGAAGCACAAAAGAAAGACACAGAGATCCCTGGGAGAAATGCCCGGCC 1569

QY 305 gccatctgggtcactgatgagcctcgccctgctgctgctgctgctgctgctgctgctgct 364

Db 1568 GCCATCTTGGGTCTCGATGACGCTCGCCCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1509

QY 365 attagaaatgaattgatgtgttctttaaaggatggcaggaagaaacagatcctgtgtg 424

Db 1508 ATTAGAAATGAATTGATGTGTCTTAAAGATGGCAGGAAACAGATCCTGTGTGG 1449

QY 425 atatttattgaacgggattacagattgaaatgaatgaatgaatgaatgaatgaatgaatga 484

Db 1448 ATATTATTGTAACGGGATTACAGATTTGAAATGAAGTCACAAAGTGAGCATTTACCAATG 1389

QY 485 agaggaacacagacagaaaaatcttgatgcttcacaagacatgcaacaaacaaatgga 544

Db 1388 AGAGGAANAACACAGAGAAAATCTTGATGGCTTCACAGACATGCANACAAACAAATGGA 1329

QY 545 atactgtgatgacatgagcagcagcaagctgggagagagataaccacgggcagaggggtca 604

Db 1328 ATACTGTGATGACATGAGCAGCCCAAGCTGGGGAGAGATAACCAGGGGCGAGGGGTCA 1269

QY 605 ggattctggccctgctgcttaaaactgctgcttcaataaccacaaatcattcatatttctaac 664

Db 1268 GGATTCTGGCCCTGCTGCTTAAACTGTGGTTTCATAAACCAAAATCATTTTCTTAAC 1209

QY 665 cctcaaaaacaaagctgttgtaatactcatctcaggtctcctctctgggcccacattct 724

Db 1208 CCTCAAAACAAAGCTGTGTGTAATATCTGATCTCTACGGTTCTCTGGGCCCAACATCT 1149

QY 725 ccatatccagccacactcatttttaattatttagtctccagatcctgtactgtgaccttt 784

Db 1148 CCATATATCCAGCCACACTCAATTTTAAATATAGTTCCACATCTGTACTGTGACCTTT 1089

QY 785 ctacactgtagaataaacattactcattttgttcaagaccctctgtgtgtgtgcctaata 844

Db 1088 CTACACTGTAGAATAACATTTACTTCATTTTGTTCAAAGAGCCCTTCGTGTGTGCTGCTTAATA 1029

Qy 845 tgaagtgactgttttttcttaaggagtgcttctgcccaggggagtgctgtgaacagctggg 904
Db 1028 tgaagtgactgttttttcttaaggagtgcttctgcccaggggagtgctgtgaacagctggg 969
Qy 905 aagcatcgaagatcttccaggggttactactactagcacagcatgatcatcagag 964
Db 968 aagcatcgaagatcttccaggggttactactactagcacagcatgatcatcagag 909
Qy 965 tgaattactaataacatcatcctcagtgcttcttgcccatactgaaatttcctccac 1024
Db 908 tgaattactaataacatcatcctcagtgcttcttgcccatactgaaatttcctccac 849
Qy 1025 ttttggccattctcaagacacctcaaatgtctattccatttaataatcacaggaattt 1084
Db 848 ttttggccattctcaagacacctcaaatgtctattccatttaataatcacaggaattt 790
Qy 1085 ttttttaacctggaagaattcaatgtttacatgcagctatggaatttaattacattt 1144
Db 789 ttttttaacctggaagaattcaatgtttacatgcagctatggaatttaattacattt 730
Qy 1145 ttttttcagtgcaagatgactaagctctttatccctccctcttctgtttgttttttc 1204
Db 729 ttttttcagtgcaagatgactaagctctttatccctccctcttctgtttgttttttc 670
Qy 1205 cagtataaagttaaaatgcttagcctgttactgaggtgtgtatcacg-cacagcctctccc 1263
Db 669 cagtataaagttaaaatgcttagcctgttactgaggtgtgtatcacg-cacagcctctccc 610
Qy 1264 catcctccagccttatctgtcatcaccaatcaacccctcccaatnysacctaaacaaatc 1323
Db 609 catcctccagccttatctgtcatcaccaatcaacccctcccaatnysacctaaacaaatc 551
Qy 1324 taacttgtaattccttgaaatgctcagncatacaatttttctctgctgagaagctct 1383
Db 550 taacttgtaattccttgaaatgctcagncatacaatttttctctgctgagaagctct 492
Qy 1384 tcttctctcttaactcagaatgctgaaagtgttgaatgaagtgactatctacttca 1443
Db 491 tcttctctcttaactcagaatgctgaaagtgttgaatgaagtgactatctacttca 432
Qy 1444 tgcaagaaggagacacatatgatgattcatcatcacatgagacagcaaatactaaaagtgt 1503
Db 431 tgcaagaaggagacacatatgatgattcatcatcacatgagacagcaaatactaaaagtgt 372
Qy 1504 aatttgattataagatttagataaaatatatgaatgaagkccacagaggggaattgtt 1563
Db 371 -atttgattataagatttagataaaatatatgaatgaagkccacagaggggaattgtt 313
Qy 1564 atggggcacgtttgtaagcctgggagtgaaqmaaggcagggaacctcatagtatctta 1623
Db 312 atggggcacgtttgtaagcctgggagtgaaqmaaggcagggaacctcatagtatctta 253
Qy 1624 tataataactctatctctatctctatcaataatcaacaaagcttttccagaattc 1683
Db 252 tataataactctatctctatctctatcaataatcaacaaagcttttccagaattc 193
Qy 1684 atgcaagtcaaatcccaagtgaaacctttatccatttgatgctgagtcgctttagaat 1743
Db 192 atgcaagtcaaatcccaagtgaaacctttatccatttgatgctgagtcgctttagaat 133
Qy 1744 tttggcaaatcatcactgggtcacttatctcaactttgagatgtgttctctgtagttaa 1803
Db 132 tttggcaaatcatcactgggtcacttatctcaactttgagatgtgttctctgtagttaa 73
Qy 1804 ttgaagaagaataggcactcttctgagccacttttaggggttccactcctggcaataagaat 1863
Db 72 ttgaagaagaataggcactcttctgagccacttttaggggttccactcctggcaataagaat 13
Qy 1864 ttacaaga 1872
Db 12 ttacaaga 4

RESULT 7
AAV62429
ID AAV62429 standard; cDNA; 820 BP.
XX
AC AAV62429;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 3.
XX
KW Prostate cancer antigen cDNA splice variant 3; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
XX
PS Claim 4; Pages 77-78; 11pp; English.
XX
CC The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 3 sequence comprising of exons 1, 3, and 4a
CC of the PCA3 gene. The PCA3 cDNA splice variant 3 sequence,
CC isolated from a human prostatic tumour tissue cDNA library,
CC was found in approximately 15% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunoassay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (antagonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 820 BP; 262 A; 169 C; 191 G; 198 T; 0 other;

Query Match 43.8%; Score 820; DB 19; Length 820;
Best Local Similarity 100.0%; Pred. No. 1.9e-225;
Matches 820; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 aagaagctgcatcagaaaaaacagaggggagatttgtgtgctgcagcgaggagaccag 60
Db 1 aagaagctgcatcagaaaaaacagaggggagatttgtgtgctgcagcgaggagaccag 60
Qy 61 gaagatctctgctggtgggaaggacctgatgatacagaggtgagaaataagaagcgtgct 120
Db 61 gaagatctctgctggtgggaaggacctgatgatacagaggtgagaaataagaagcgtgct 120
Qy 121 gactttaccatctgagggccacacatctctgtaaatggagataattaacatcactgaaac 180
Db 121 gactttaccatctgagggccacacatctctgtaaatggagataattaacatcactgaaac 180


```

XX Homo sapiens.
OS WO200004149-A2.
PN 27-JAN-2000.
PD 14-JUL-1999; 99WO-US15838.
PF 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 15-JAN-1999; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX (CORI-) CORIXA CORP.
PA Dillion DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX WPI; 2000-171268/15.
XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein.
XX Claim 1; Page 199-200; 263pp; English.
XX The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.
XX Sequence 718 BP; 222 A; 145 C; 169 G; 172 T; 10 other;
SQ
Query Match 27.4%; Score 513.2; DB 21; Length 718;
Best Local Similarity 97.3%; Pred No. 2.3e-137;
Matches 585; Conservative 0; Mismatches 9; Indels 7; Gaps 6;
QY 27 ggagatttggtgg-ctgcagccgaggagaccaggaagatctgcatgtgtggaagacc 85
Db 1 ggagatttggtgttcagccgaggagaccaggaagatctgcatgtgtggaagacc 60
QY 86 tgatgatacagaggtagaataaagaagctgctgactttaccatctgagccacacat 145
Db 61 tgatgatacagaggtagaataaagaagctgctgactttaccatctgagccacacat 120
QY 146 ctgctgaatggagataataacatcactagaaacagcaagatgacataatgctctaa 205
Db 121 ctgctgaatggagataataacatcactagaaacagcaagatgacataatgctctaa 180
QY 206 gtatgacatgttttgcacatttcagccctttaaataacacacacaggaagcac 265
Db 181 gtatgacatgttttgcacatttcagccctttaaataacacacacaggaagcac 240
QY 266 aaaagaacacagagatccctgggagaatacccgccgcatcttgggtcatgatga 325
Db 241 aaaagaacacagagatccctgggagaatacccgccgcatcttgggtcatgatga 300
QY 326 gctcgcctctgctcctggtccgctgtgaggaagacattagaaaaatgaattgatgtg 385
Db 301 gctcgcctctgctcctggtccgctgtgaggaagacattagaaaaatgaattgatgtg 360

```

QY 386 ttccctaaaggatggcaggaagaaacagatccctgtgtgtgatatattttgaacgggatta 445
 Db 361 ttccctaaaggat-ggcaggaagaaacagatccctgtgtgtgatatattttgaacgggatta 419
 QY 446 cagatttgaatgaagtcaaaagtgcacattaccatgagaggaagaaacagacagagaaa 505
 Db 420 cagatttgaatgaagtcaaaagtgcacattaccatgagaggaagaaacagacagagaaa 479
 QY 506 tcttgatgcttcacagacatgcaacaaacaaatggaatactgtgatgacatgaggca 565
 Db 480 tcttgatgg-ttcacagacatgcaacaaacaaatggaatactgtgatgacacagag--c 536
 QY 566 gccagctgggagagagataaacacgggcagagggctcaggattctggccctgctccta 625
 Db 537 agccaactgggagagagat-accacggggcaga-ggtcaggattctggccctgctccta 594
 QY 626 a 626
 Db 595 a 595
 RESULT 11
 AAC06768
 ID AAC06768 standard; cDNA; 437 BP.
 XX AAC06768;
 AC AAC06768;
 XX 06-OCT-2000 (first entry)
 DT Human secreted protein 5' EST, SEQ ID NO: 10843.
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 XX gene therapy; chromosome mapping; ss.
 KW Homo sapiens.
 OS EP1033401-A2.
 PN 06-SEP-2000.
 PD 21-FEB-2000; 2000EP-0200610.
 XX 26-FEB-1999; 99US-0122487.
 PR (GEST) GENSET.
 PA Dumas Milne Edwards J, Duclert A, Giordano J;
 PI WPI; 2000-500381/45.
 DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX Claim 1; SEQ ID 10843; 71pp + CD-ROM; English.
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX Sequence 437 BP; 140 A; 87 C; 118 G; 88 T; 4 other;

Query Match 22.1%; Score 414.2; DB 21; Length 437;
 Best Local Similarity 99.0%; Pred. No. 4.8e-109;
 Matches 413; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 agaagctggcatcagaaaaacagagggagattgtgtgctgcagccgagggagaccag 60
 DB 21 agaagctggcatcagaaaaacagagggagattgtgtgctgcagccgagggagaccag 80
 QY 61 gaagatctcatgtgtgggaagacacctgatgatacagagtgagaaataagaaggctgct 120
 DB 81 gaagatctcatgtgtgggaagacacctgatgatacagagtgagaaataagaaggctgct 140
 QY 121 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaac 180
 DB 141 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaac 200
 QY 181 agcaagatgacaataatgtctaagtagtgacatgtttttgcacatttccagcccttt 240
 DB 201 agcaagatgacaataatgtctaagtagtgacatgtttttgcacatttccagcccttt 260
 QY 241 aaatatccacacacagagaaacacaaaaggaagcacagatccctgggaaatgcc 300
 DB 261 aaatatccacacacagagaaacacaaaaggaagcacagatccctgggaaatgcc 320
 QY 301 gcccgcacatctgggtcatcgatgagcctgcgcctgtgctgctgctgctgagggaa 360
 DB 321 gcccgcacatctgggtcatcgatgagcctgcgcctgtgctgctgctgagggaa 380
 QY 361 gacattagaaaatgaattgattgttctttaaagatgggaggaacacagatcct 417
 DB 381 gacattagaaaatgratgtgttctttaaagatgggaggaacacagatcct 437

RESULT 12
 AAZ33445
 ID AAZ33445 standard; cDNA; 359 BP.
 XX
 AC AAZ33445;
 XX
 DT 08-DEC-1999 (first entry)
 XX
 DE Human prostate cancer-associated EST 23.
 XX
 KW Expressed sequence tag; EST; prostate tumor; antitumor; treatment;
 KW gene therapy; tissue specificity human; ss.
 XX
 OS Homo sapiens.
 XX
 PN DE19811193-A1.
 XX
 PD 16-SEP-1999.
 XX
 PF 10-MAR-1998; 98DE-1011193.
 XX
 PR 10-MAR-1998; 98DE-1011193.
 XX
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX
 PI Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosenthal A;
 XX
 DR WPI; 1999-519628/44.
 DR P-PSDB; AAY48243.
 XX
 PT New nucleic acid expressed at high level in prostatic tumor tissue and
 PT encoded polypeptides, useful for treating cancer and screening for
 PT therapeutic agents
 XX
 PS Claim 1a; 87; 166pp; German.
 XX
 CC This invention describes novel nucleic acid sequences (A) that are
 CC expressed at high level in prostatic tumor tissue and encode gene
 CC products or their fragments. The products of the invention have

CC antitumor activity. Polypeptides (I) encoded by (A) are used: (i) for
 CC identifying agents for treatment of prostatic cancer and (ii) for
 CC therapy of prostate cancer. Optionally where expressed by gene therapy
 CC methods. (A) is also used to isolate full-length genes (for gene therapy)
 CC and for recombinant production of (I), which can be used to raise
 CC specific antibodies. (A) are identified by assembly of ESTs (expressed
 CC sequence tags) before they are analyzed for expression pattern (tissue
 CC specificity). This approach eliminates many of the false results, as
 CC regards tissue specificity, associated with known methods that use
 CC single (usually short) ESTs. AAZ33423-233476 represent expressed
 CC sequence tags described in the method of the invention.

XX Sequence 359 BP; 121 A; 75 C; 94 G; 69 T; 0 other;

Query Match 17.6%; Score 330; DB 20; Length 359;
 Best Local Similarity 100.0%; Pred. No. 6.5e-85;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaagctggcatcagaaaaacagagggagattgtgtgctgcagccgagggagaccag 60
 DB 30 agaagctggcatcagaaaaacagagggagattgtgtgctgcagccgagggagaccag 89
 QY 61 gaagatctcatgtgtgggaagacacctgatgatacagagtgagaaataagaaggctgct 120
 DB 90 gaagatctcatgtgtgggaagacacctgatgatacagagtgagaaataagaaggctgct 149
 QY 121 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaac 180
 DB 150 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaac 209
 QY 181 agcaagatgacaataatgtctaagtagtgacatgtttttgcacatttccagcccttt 240
 DB 210 agcaagatgacaataatgtctaagtagtgacatgtttttgcacatttccagcccttt 269
 QY 241 aaatatccacacacagagaaacacaaaaggaagcacagatccctgggagaaatgcc 300
 DB 270 aaatatccacacacagagaaacacaaaaggaagcacagatccctgggagaaatgcc 329
 QY 301 gcccgcacatctgggtcatcgatgagcctc 330
 DB 330 gcccgcacatctgggtcatcgatgagcctc 359

RESULT 13
 AAA06520/c
 ID AAA06520 standard; cDNA; 301 BP.
 XX
 AC AAA06520;
 XX
 DT 13-JUN-2000 (first entry)
 XX
 DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:287.
 XX
 KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
 KW immunogenic; cytostatic; vaccine; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200004149-A2.
 XX
 PD 27-JAN-2000.
 XX
 PF 14-JUL-1999; 99WO-US15838.
 XX
 PR 14-JUL-1998; 98US-0115453.
 PR 14-JUL-1998; 98US-0116134.
 PR 23-SEP-1998; 98US-0159812.
 PR 23-SEP-1998; 98US-0159822.
 PR 15-JAN-1999; 99US-0232149.
 PR 15-JAN-1999; 99US-0232880.
 PR 09-APR-1999; 99US-0288946.
 XX

PA	(CORI-) CORIXA CORP.
XX	
PI	Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX	
DR	WPI; 2000-171268/15.
XX	
PT	New polypeptide useful for treating and diagnosing prostate cancer
PT	comprises an immunogenic portion of prostate tumor protein -
XX	
PS	Claim 1; Page 192; 263pp; English.
XX	
CC	The present invention describes isolated polypeptides, comprising an
CC	immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC	and polynucleotides encoding them have cytostatic activity and can be
CC	used in vaccines and in gene therapy. The polypeptides and
CC	polynucleotides encoding them, antigen presenting cells which express
CC	the polypeptides, antibodies against the polypeptides and vaccines
CC	comprising them can be used for inhibiting the development of prostate
CC	cancer in a patient. The polypeptides can be used to generate antibodies
CC	or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC	the polynucleotides encoding the polypeptides can be used as a probe or
CC	to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC	AAA82000 to AAA82020 represent sequences used in the exemplification of
CC	the present invention.
XX	
SQ	Sequence 301 BP; 76 A; 58 C; 70 G; 97 T; 0 other;
Query Match 15.4%; Score 288.4; DB 21; Length 301;	
Best Local Similarity 99.3%; Pred. No. 5.3e-73;	
Matches 300; Conservative 0; Mismatches 1; Indels 1; Gaps 1	
QY	472 agcattaccatgagaggaaaacagacgagaaaaattctgatgcgttcacaagacatgcaa 531
DB	301 AGCATATACCATTGAGGGAAAAACAGACGAGAAAACTTGTGCGTTCAACAAGACATGCCAA 242
QY	532 caaacaaatacggaactactgtgatgcacatgaggcagccaagtggggaggagataaccacg 591
DB	241 CAACAAAAATGGAATACTGTGTAACATGAGGCAGCCAAGCTGGGAGGAGATAACCACG 182
QY	592 gggcagagggtcaggattctggccctgtgcctaaaactgtgcgttcataaaccaaatcatt 651
DB	181 GGGCAGAGGGTCAGGATTCTGGCCCCCTGCCTAAACTGTCGTTTCATAAACCAATCAT 122
QY	652 tcatatttctaaccctcaaaacaaaagtgttgttaatatctgcacctcactgcgttccttg 711
DB	121 TCATATTTCTAACCTTCAAAACAAACACTGTTGTAATATCTGATCTAC-GTTCCTTCG 63
QY	712 ggcccacaattctccatatatccagccacactcatttttaatattagttccccagatctg 771
DB	62 GGCCCAACATTTCCATATATCCAGCCACACTCATTTTTTAATATTAGTTCOCGATCTG 3
QY	772 ta 773
DB	2 TA 1
RESULT 14	
AAP58252/c	
ID	AAF58252 standard; DNA; 936 BP.
XX	
AC	AAF58252;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D1835.
XX	
KX	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
XX	
OS	Synthetic.
PN	WO200107665-A2.

Qy 1567 gggcagctttagcctgggatggaagmaaaggcagggaacccctcatagttattat 1626
Db 177 WWWWWW... 118
Qy 1627 aatactacttctctatctctatcaataatccaagaagcttttccacagaattcatg 1686
Db 117 WWWWWW... 58
Qy 1687 cagtgcacatcccaaggtaacctttatccatttcattggtgagtcgctttagaat 1743
Db 57 WWWWWW... 1

RESULT 15
AAF58254/C
ID AAF58254 standard; DNA; 936 BP.
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX

PT Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX

CC The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 7.3%; Score 136; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. NO. 5.9e-29;
Matches 7; Conservative 491; Mismatches 279; Indels 0; Gaps 0;

Qy 967 aattatctaatacatcatcctcagtgctgtcttggccatctactgaataattccactt 1026
Db 777 WWWWWW... 718
Qy 1027 ttgtgccattctcaagacctcaaaatgtcattccatttaataatcacaggattattt 1086
Db 717 WWWWWW... 658
Qy 1087 ttttaacctggaagaattcaatgttacatgcagctatgggaatttaattacatatattg 1146

Db 657 WWWWWW... 598
Qy 1147 ttltccagtcgaaagatgactaaagtcctttatccctcctctgtttgtattttttcca 1206
Db 597 WWWWWW... 538
Qy 1207 gtataaagttaaaatgcttagcttactgaggtgtatatacagcacagcctctcccat 1266
Db 537 WWWWWW... 478
Qy 1267 cctccagccttatctgtcatcaccatcaacccctcccatnysacctaaacaaatctaa 1326
Db 477 WWWWWW... 418
Qy 1327 cttgtaattccttgacatgctcagncatacatatttctctcgtcgtgagaagctctcc 1386
Db 417 WWWWWW... 358
Qy 1387 ttgtctcttaantctagaatgatgtaaagttttgataagttgacttatcttactcatgc 1446
Db 357 WWWWWW... 298
Qy 1447 aaagaaggacacatagattcatcatcacatgagacagcaataactactaaagttaat 1506
Db 297 WWWWWW... 238
Qy 1507 ttgattataagagtttagataaataatgaatgaagakccacagaggggaatgttatg 1566
Db 237 WWWWWW... 178
Qy 1567 gggcagctttagcctgggatggaagaaaggcagggaacccctcatagttattat 1626
Db 177 WWWWWW... 118
Qy 1627 aatactacttctctatctctatcaataatccaagaagcttttccacagaattcatg 1686
Db 117 WWWWWW... 58
Qy 1687 cagtgcacatcccaaggtaacctttatccatttcattggtgagtcgctttagaat 1743
Db 57 WWWWWW... 1

Search completed: July 31, 2001, 07:26.09
Job time: 17781 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:41:09 ; Search time 10902.7 Seconds
(without alignments)
2655.824 Million cell updates/sec

Title: US-09-402-713A-3
Perfect score: 1872
Sequence: 1 agaagctggcatcagaaaaa.....caataaagaatttacaaga 1872

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

72: gb_htg14.*	73: gb_htg14.*	74: gb_htg15.*	75: gb_htg16.*	76: gb_htg17.*	77: gb_htg18.*	78: gb_htg19.*	79: gb_htg20.*	80: gb_htg21.*	81: gb_htg22.*	82: gb_htg23.*	83: gb_htg24.*	84: gb_htg25.*	85: gb_pr1.*	86: gb_pr2.*	87: gb_pr3.*	88: gb_pr4.*	89: gb_pr5.*	90: gb_pr6.*	91: gb_pr7.*	92: gb_pr8.*	93: gb_pr9.*	94: gb_ro1.*	95: gb_ro2.*	96: gb_in4.*	97: gb_pr10.*	98: em_ba3.*
----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	---------------	--------------

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	
1	1772.4	94.7	3923	88	AF103907 Homo sapi	
2	1589.4	84.9	5435	88	AF103908 Homo sapi	
3	1555.8	83.1	164371	80	AL390239 Homo sapi	
4	1361.4	72.7	173831	80	AL359314 Homo sapi	
c	5	1317.8	70.4	267581	80	AL358573 Homo sapi
6	330	17.6	359	9	AX018075 Sequence	
7	100.2	5.4	143675	79	AL161625 Homo sapi	
8	100.2	5.4	267581	80	AL358573 Homo sapi	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Db	2036	TGCCTGGTCCCGCTTGAGGGAGACATTTAGAAAAATGAATTTGATGTGTTTCCCTTAAAGG	2099
Qy	397	atggcaggagaaacagatcctgtgtggatatatttgaacgggattacagatttgaaaa	456
Db	2096	ATGGCGAGGAAAAACAGATCCTGTGTGGATATTTATTTGAACGGGATTACAGATTTGAAA	2155
Qy	457	tgaagtcaaaaagtgcagctattaccaatgagaggaaaaacagacgagaaaaattcttgatggct	516
Db	2156	TGAAGTGCACAAAGTGAGCATTTACCAATGAGAGGAAAAACAGACGAGAAAACTTTGATGGCT	2215
Qy	517	tcaagagacatgcaacaacaaaaatggaaactgtgatgacatgagcgcccaagctggg	576
Db	2216	TCACAGACATGCAACAAACAAAATGGAATACTGTGATGCATCAGGACGCCAAGCTGGG	2275
Qy	577	gaggagataaacacggggcagagggtcaggattctggccctgcctgcctcctaaactgtgcgtt	636
Db	2276	GAGGAGATAACACGGGGCAGAGGTCAGGATTCGGGCCCTGCTGCCCTAAACTGTGGCTT	2335
Qy	637	catacccaaatcatcttcataatttctaaccctcctaaacaaagctgttgtaatatctgatct	696
Db	2336	CATAACCAAAATCATTTTCATATTTCAACCCTCAAAACAAAGCTGTTGTAATATCTGATCT	2395
Qy	697	ctacggcttctctggggcccaacattctccatatatccagcccaacactcaattitaatat	756
Db	2396	CTACGGTTCCTCTGGGCCCAACATCTCCATATATCCAGCCACACACTCAATTTTAATATT	2455
Qy	757	tagtctccagatctgactgtgaccttctcacactgtagaataacattactcatttggtt	816
Db	2456	TAGTTCCCAGATCTGTACTGTGACCTTTCTACCTGTAGAAATACATTTACTAATTTGTT	2515
Qy	817	caagacccttggtgtgctgctctaattgttagctgactgttttctcaaggagtgctct	876
Db	2516	CAAAAGCCCTCTGTTGTGTGCTTAATATGTAGCTGACTGTTTCTCCTAAGGAGTGTTCT	2575
Qy	877	ggcccgagggaatgtgaaacaggctgggaagcatcacaagatcttccagggttatactt	936
Db	2576	GGCCCCAGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGGGTATACTT	2635
Qy	937	actagcacacagatgatcattcacggagtgaattatctaaatcaacatcatctccagtgct	996
Db	2636	ACTAGCACACAGCTGATCTTACGGAGTGAAATTTATCTAATCAACATCATCTCTCAGTGTC	2695
Qy	997	tttgcccaactactgaaattcatttcccacttttgtgcccattctcaagaccctcaaaatgctc	1056
Db	2696	TTTGGCCATATCGAAATTCATTTCCACTTTTGTGGCCATCTCTCAAGACCTCAAAATGTC	2755
Qy	1057	attccaattaatcacagattaaacttttttttaacctggagaattcaatgtttacat	1116
Db	2756	ATTCCATTAATATCACAGGATTAAGCTTTTTTTTTTTTAACTCGGAAGAATTCAATGTGTACAT	2815
Qy	1117	gcagctatgggaattaaattacatatattgttttccagtgcaagatgacaaagtccttt	1176
Db	2816	GCAGCTATGGGAATTTAATTTACATATTTTGTGTTTCCAGTGCAAAAGATGACTAAGTCTCTT	2875
Qy	1177	atccctcccccttggttgatttttttccagttataaagttaaaatgccttagccttgtaact	1236
Db	2876	ATCCCTCCCTTTGTTTGATTTTTTTTTTCCAGTATAAAGTTAAAAATGCTTAGCCTTGACT	2935
Qy	1237	gaggctgtatacagacagcctctcccatccctccagccttatctgtcatcaccaatcaa	1296
Db	2936	GAGGCTGTATACAGCACAGCCTCTCCCATCCCTCCAGCCCTTATCTGTCTCATCACCATCAA	2995
Qy	1297	ccccctcccatnysacctaaacaaaaatcaacttgttaattcctgaaacatcatcaggncata	1356
Db	2996	CCCCCTCCCATACACCTTAACAAAAATCTAATCTTGTAATCTCTGAAATGTCAGGACATA	3055
Qy	1357	catrttccctctgcctgagaagctcttccctgtctctcttaantctagaatgagttaagt	1416
Db	3056	CATTATTCTCTGCTGTGAGAGCTCTTCCTTGCTCTTAAATCTAGAAATGATGAAGT	3115
Qy	1417	tttgaataagttgactatatcttacttcatacgaaagaagggaacacatacgagattcatc	1476
Db	3116	TTTGAATAAGTTGACTATCTTACTTCATGCAAGAGGGACACATATGACATTCATCATC	3175


```
|||||
Db 34874 TTAGGTTTCACTCTCGCAATAAGAATTACAAAGA 34910
|||||
RESULT 4
AL359314 173831 bp DNA HTG 08-APR-2001
LOCUS Homo sapiens chromosome 9 clone RP11-108L4, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.
ACCESSION AL359314
VERSION AL359314.12 GI:13396560
KEYWORDS HTG; HTGS-PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173831)
Leongamornlert,D.
Direct Submission
Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:13277120.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bA108L4
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 173545 bases at least Q40
Consensus quality: 173599 bases at least Q30
Consensus quality: 173621 bases at least Q20
Insert size: 173631; sum-of-contigs
Insert size: 172123; 10.0% error; agarose-fp
Quality coverage: 10.35x in Q20 bases; sum-of-contigs Quality
coverage: 10.50x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 50595: contig of 50595 bp in length
* 50596 50695: gap of 100 bp
* 50696 157636: contig of 106941 bp in length
* 157637 157736: gap of 100 bp
* 157737 173831: contig of 16095 bp in length.
FEATURES
Location/Qualifiers
1..173831
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-108L4"
/clone_lib="RPC1-11.1"
/1..50595
/2..157636
/3..157736
/4..173831
/misc_feature
1..50595
/1..50595
/misc_feature
157636..157636
/157636..157636
/misc_feature
157737..173831
/157737..173831
BASE COUNT 52107 a 36469 c 35018 g 50037 t 200 others
ORIGIN
```

```
Query Match 72.7%; Score 1361.4; DB 80; Length 173831;
Best Local Similarity 92.4%; Pred. No. 0;
Matches 1395; Conservative 4; Mismatches 107; Indels 3; Gaps 3;
QY 277 cagagatccctgggagaaatgccgcgcgcgcctcttggtgcatcatgatgagcctgcgcctg 336
|||
Db 156237 CACAGATCCCTGGGAGAAATGCCGCGCCCATCTTGGGTCTCATCGATGAGCCTCGCCCTG 156296
QY 337 tgcctggtccctctgtgagggagagacattagaaaaatgaatgatgtgttctcttaagg 396
|||
Db 156297 TCCTCTGTCCTCCCTTGTGAGGGAAGACATTAAGAAATGAATGTCTTCTTAAAGG 156356
QY 397 atgggcaggaacacagatcctgtgtgggataattttgaacgggattacagatttgaa 456
|||
Db 156357 ATGGCAGGAAACACAGATCTCTGTGTGGGATATTTATTTGACGGGATACAGATTGAAA 156416
QY 457 tgaagtcaacaagtgcattaccatggagagagagagagagagagagagagagagagag 516
|||
Db 156417 TGAAGTCAAAAAGTGAGCATTTACCAATGAGAGGAAACAGACGAGAGAAATCTTGATGGCT 156476
QY 517 tcacaagacatcaacaacaaataaggaatactgtgatactgatactgaggaagcagcaagctgg 576
|||
Db 156477 TCACAAGACATGCAACAAACAAATGGAATCTGTGATGACATGAGGACGCAAGCTGGG 156536
QY 577 gaggagataaacacaggggcagaggggtcaggattctctggccctgctgctaaactgtgcgtt 636
|||
Db 156537 GAGGAGATAAACCCGCGGCGAGAGGTCAGGATTCGGCCCTGCTGCCTAAACTGTGCGTT 156596
QY 637 cataaccaatcatcttcatatttctaaccctcaaaacaaagctgttgttaatactgatct 696
|||
Db 156597 CATAACCAAAATCATTTATATTTCTTAACCTCAAAACAAAGCTGTGTGTAATATCTCATCT 156656
QY 697 ctacggttctctgggcccacattctccatatactatccagccacacactcatttttaatt 756
|||
Db 156657 CTACGGTTCCTTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCATTTTATATTT 156716
QY 757 tagttccagatctgtactgtgacctttctacactgtagaataaacattactcttttgtt 816
|||
Db 156717 TAGTTCCCGAGATCTGTACTGTGACCTTTCTACACTGTAGAATAACATTTACTCATTTGTT 156776
QY 817 caaagacctctgtgtgctgctctaataatgtagctgactgttttctcctcaagagtgctt 876
|||
Db 156777 CAAAGACCTTCGTGTGTGCTGCTAATATGTAGTACTGCTGTTTCTTCTAAGGAGTGTCT 156836
QY 877 gccccagggatctgtgaacagcctgggaagcatctcaagatcttccaggttatactt 936
|||
Db 156837 GCCCAGGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGCTTATACCT 156896
QY 937 actagcacacagcatgatacttaccgagtgagtgaaattatctaatcaacatcatcctcagtgct 996
|||
Db 156897 ACTAGCACACAGCATGATCATTTACGAGTGGAATTTATCTAATCAACATCATCTCAGTGTC 156956
QY 997 ttgcccatactgaaatcattcccaacttttggcccaacttccagctcaggaagcctcaaatgtc 1056
|||
Db 156957 TTTGGCCATCTACTGAAATTCATTTCCACTTTTGTGCCCATCTTCAAGAGCTTCAAAATGTC 157016
QY 1057 attccattaatatcacagattaaacttttttttaacctggaagaattcaatgtttacat 1116
|||
Db 157017 ATTCCATTAAATACAGGATTAACCTTTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 157076
QY 1117 gcagctatgggaatttaatacatattttgttttccagtgcaaaagatgactcaagtccttt 1176
|||
Db 157077 GCAGCTATGGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 157136
QY 1177 atccctccctctgtgttgatttttttccagtaataaagttaaaatgcttagcctgtact 1236
|||
Db 157137 ATCCCTCCCTTTGTGTGATTTTTCAGATATAAAGTTAAATGCTTTAGCTTTGTACT 157196
QY 1237 gaggctgtatcacag-cacagcctctcccatccctccagccttattctgtcatcacatca 1295
|||
Db 157197 GAGGCTGTATACAGCCACAGCCTCTCCCATCTCCCTCCAGCCTTATCTGTCTCATCACCATCA 157256
```



```
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.
```

FEATURES

Location/Qualifiers

```
1..267581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-133022"
/clone.lib="RPC1-11.1"
1..25718
/note="assembly_fragment:00645"
fragment_chain:1
clone_end:77
vector_side:left
25819..29567
/note="assembly_fragment:01660"
fragment_chain:1
29668..40520
/note="assembly_fragment:03080"
fragment_chain:1
40621..50002
/note="assembly_fragment:00771"
fragment_chain:1
50103..54942
/note="assembly_fragment:02629"
fragment_chain:1
55043..59067
/note="assembly_fragment:04565"
fragment_chain:1
59168..72327
/note="assembly_fragment:00223"
fragment_chain:1
72428..79396
/note="assembly_fragment:03318"
fragment_chain:1
79497..82738
/note="assembly_fragment:00224"
fragment_chain:2
82839..87543
/note="assembly_fragment:03050"
fragment_chain:2
87644..89847
/note="assembly_fragment:01948"
fragment_chain:3
89948..92376
/note="assembly_fragment:02191"
fragment_chain:3
92477..99799
/note="assembly_fragment:00010"
99900..102117
/note="assembly_fragment:00672"
102218..109330
/note="assembly_fragment:01326"
109431..112187
/note="assembly_fragment:01498"
112288..116407
/note="assembly_fragment:01607"
116508..119386
/note="assembly_fragment:01899"
119487..121869
/note="assembly_fragment:02068"
121970..125546
/note="assembly_fragment:02469"
125647..128990
/note="assembly_fragment:02502"
```

```
misc_feature 129091..134055
/note="assembly_fragment:02669"
misc_feature 134156..138314
/note="assembly_fragment:02757"
misc_feature 138415..140612
/note="assembly_fragment:02776"
misc_feature 140713..143296
/note="assembly_fragment:03127"
misc_feature 143397..148113
/note="assembly_fragment:03634"
misc_feature 148214..151009
/note="assembly_fragment:03898"
misc_feature 151110..156616
/note="assembly_fragment:04220"
misc_feature 156717..158763
/note="assembly_fragment:04276"
misc_feature 158864..162129
/note="assembly_fragment:04418"
misc_feature 162230..164443
/note="assembly_fragment:04428"
```

Query Match 70.4%; Score 1317.8; DB 80; Length 267581;

Best Local Similarity 91.9%; Pred. No. 0;

Matches 1355; Conservative 4; Mismatches 113; Indels 3; Gaps 3;

```
QY 277 cagagatccctggagaaatgccgcgcctcttgggtcatcgatgagcctcgccctg 336
|||
Db 165900 CACAGATCCCTGGGAGAAATGCCGCCGCCCATCTTGGGTCTCGATGAGCCTGCCCTG 165841

QY 337 tgccctggctcccgctgtgaggaagacattagaaaaatgaattgatgttccctaaagg 396
|||
Db 165840 TGCCTGGTCCCGCTTGTGAGGGAAGACATTAGAAAATGAATGTGTCTTCTTAAAGG 165781

QY 397 atgggcaggaataacagatcctgttggatatatttgaacgggattcacagatttgaaa 456
|||
Db 165780 ATGGGCAGGAATAACAGATCCTGTGTGGATATTTATTTCACGGGATTCACAGATTGAAA 165721

QY 457 tgaagtccacaaagtgcattaccatgcagaggaagaaacagacagaaaaatcttgaaggct 516
|||
Db 165720 TCAAGTCACAAAGTGAGCATTACCAATGAGAGGNAACAGACAGAGAAATCTTGATGGCT 165661

QY 517 tcacaagacatgcacacaaacaaatggaatactgtgatatgacatgagggcagcaagctggg 576
|||
Db 165660 TCACAAGACATGCAACAAACAAATGGAATACTGTGATGACATGAGCGACGCAAGCTGGG 165601

QY 577 gagagataaacacagggcgagaggtcaggattcttggccctgcctcctaaactgtcgctt 636
|||
Db 165600 GAGGAGATAACCCAGGCGAGAGGGTCAGGATCTGGCCCTGCTGCCTAAACTGTGCGTT 165541

QY 637 catacacaatcatcttcatatttctaaccctcaaaacaaagctgtgtgtaatatctgatct 696
|||
Db 165540 CATAACCAATCATTTTCATATTTCTAAACCTCAAAACAAAGCTGTGTGTAATATCTGATCT 165481

QY 697 ctacaggttccctctgggcccacattctccatatatccagccacacactatcttataatt 756
|||
Db 165480 CTACGGTTCCTTCTGGGCCCAACATTTCTCCATATATCCAGCCACACATCAATTTTAAAT 165421

QY 757 tagttccagatctatctactgtgacctttctcacagtgaataaacattactctttgtt 816
|||
Db 165420 TAGTTCACAGATCTGTACTGTGACCTTCTACACGTAGAAATCAACATTAATCTCATTTGTT 165361

QY 817 caaagaccttcgtgtgctgctaataatgtagctgactgttttttctcaaggagtgcttct 876
|||
Db 165360 CAAAGACCTTCGTGTGTGCTGCTAATATGTAGCTGACTGTTTTCCTTAAGGAGTGTCT 165301

QY 877 gcccaggggattctgtgaacagggctgggaagcatctcaagatctttccagggttatactt 936
|||
Db 165300 GCCCCAGGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGTTTATCTT 165241

QY 937 actaacacacagatgattacagagtggaattatctaatcaacatcatctcagtgctc 996
|||
Db 165240 ACTAGCACACAGATGATCATTTACGGAGTGAATATATCTAATCAACATCATCTCAGTGTC 165181
```

```
Qy 997 ttgtcccaactgaattccattcccaacttttgccttccatttgccttcccaactcctcaaatgtc 1056
Db 165180 TTTGCCCATACTGAAATTCATTCCCACTTTGTGCTCCCACTTCTCAAGACCTCAAAATGTC 165121
Qy 1057 attccattaatatcacagattaaacttttttttttttttttttttttttttttttttttttttt 1116
Db 165120 ATTCCATTAAATATCACAGGATTAACCTTTTFTTTTAACTGGAAGAAATTCATGTTTACAT 165061
Qy 1117 gcagctatgggaatttaattacatatattgttttccagtgcaaatgactaaagtccttt 1176
Db 165060 GCAGCTATGGGAATTAATACATATTTGTTTCCAGTGCAAGAGATGACTAAAGTCCCTTT 165001
Qy 1177 atccctccccctttgttgatttttttccagttataaagttaaagtcttagccttgact 1236
Db 165000 ATCCCTCCCCCTTTGTTGATTTTGTGTTTCCAGTATAAAGTTAAATGCTTAGCCTTGACT 164941
Qy 1237 gaggctgtatacag-cacagcctctccccatccccctccagccttatctgtcatcacatca 1295
Db 164940 GAGGCTGTATACAGCCACAGCCTCTCCCACTCCCTCCAGCCTTATCTGTCAATCACATCA 164881
Qy 1296 accctcccatnysaccataaaactaaacttaacttgaattcccttgaacatgtcagpnat 1355
Db 164880 ACCCTCCCATG-CACCTAAACAAATCTAAGTGTGTAATCTCTGCAACATGTCAGG-CAT 164823
Qy 1356 acattttccctctgcctgagaagctcttcttcttcttcttaanttagaatgatgataag 1415
Db 164822 ACATTATTCTTCTGCTGAGAAGCTCTTCCCTTGTCTCTTAAATCTAGAATGATGATAAG 164763
Qy 1416 ttttgaataagtgtactatttacttccatgcaaaagaaggacacatatagattcattcat 1475
Db 164762 TTTTGAATAAGTGTACTTCTTACTTCTATGCAAAAGAGGACACATATGAGATTCTATCAT 164703
Qy 1476 cacatgacacagcaataactaaagtgaattgtattgaagtttagataaataatgtg 1535
Db 164702 CACATGACACAGCAAACTAAAGTGAATTTGATTAAGAGTTTAGATAAATAATAG 164643
Qy 1536 aaatgaagakccacagaggggaatttttatggggcagctttgtgaagcctgggatgtgaag 1595
Db 164642 AAATGCAAGAGCCACAGAGGAATGTTTATGGGCACGCTTTGTAAGCCTGGGATGTGAAG 164583
Qy 1596 maaagcagggaacctatgattatttatataataataacttcttctctatctctatcac 1655
Db 164582 CAAAGCGAGGGAACCTCATGATCTTATATATATATTTNNNNNNNNNNNNNNNNNNNNNN 164523
Qy 1656 aatatcaacaagctttcacagaattcatgcagtgcaaaatcccccaaggtaacctttat 1715
Db 164522 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 164463
Qy 1716 ccatttcattggtgagtcgctttagaattttggca 1750
Db 164462 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 164428

RESULT 6
LOCUS AX018075 359 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 23 from Patent WO9946374.
ACCESSION AX018075
VERSION AX018075.1 GI:10042526
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 359)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences from prostate tumour tissue
JOURNAL Patent: WO 9946374-A 23 16-SEP-1999;
SCHWITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
```

```
source 1. 359
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 121 a 75 c 94 g 69 t
ORIGIN
Query Match 17.6%; Score 330; DB 9; Length 359;
Best Local Similarity 100.0%; Pred. No. 4.6e-71;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 agaaactgcatcagaataaaacagaggagatttctgtggtgctgagggagagaccag 60
Db 30 AGAAGCTGGCATCAGAAAAACAGAGGGGAGATTCTGTGCTGCGAGCCGAGGAGACCAG 89
Qy 61 gaagatctgcattggaagacctgatgatacagagtgagaataagaagcgtct 120
Db 90 GAAGATCTGCTATGGTGGGAAGACCTGATGATACAGAGTGAGAATAAGAANGCTGCT 149
Qy 121 gactttaccattgagggcacacatctgctgaatggagataatgaacatcactagaac 180
Db 150 GACTTTACCATCTGAGGCCACACATCTGCTGAATGGAGATAATTAAACATCACTAGAAC 209
Qy 181 agcaagatgacaataatagtctaaagttagtgacatgtttttgcacatttccagcccttt 240
Db 210 AGCAAGATGACAATAATAGTCTAAGTAGTAGACATGTTTTCACATTTCCAGCCCTTT 269
Qy 241 aaatatccacacacagaaagacacaaaggaagacacagagatccctggagaaatgcc 300
Db 270 AAATATCCACACACAGAGAACGACAAAGGAAGACAGAGATCCCTGGGAGAAATGCC 329
Qy 301 ggcgcgcctcttgggtcatcgtatgagcctc 330
Db 330 GCCCGCCATCTGGGTCATCGATGAGCCTC 359

RESULT 7
LOCUS AL161625 143675 bp DNA HTG 20-JAN-2001
DEFINITION Homo sapiens chromosome 9 clone RP11-146P9, *** SEQUENCING IN
PROGRESS ***, 10 unordered pieces.
ACCESSION AL161625
VERSION AL161625.6 GI:9863607
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 143675)
AUTHORS Plumb,B.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:8894260.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bal46p9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 138647 bases at least Q40
Consensus quality: 140619 bases at least Q30
Consensus quality: 141706 bases at least Q20
Insert size: 142775; sum-of-contigs
Insert size: 147523; 5.9% error; agarose-fp
Quality coverage: 4.29x in Q20 bases; sum-of-contigs Quality
coverage: 4.22x in Q20 bases; agarose-fp
```


* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 12163: contig of 12163 bp in length
12164 12263: gap of 100 bp
12264 25269: contig of 13006 bp in length
25270 25369: gap of 100 bp
25370 57709: contig of 32340 bp in length
57710 57809: gap of 100 bp
57810 66792: contig of 8983 bp in length
66793 6892: gap of 100 bp
6893 102772: contig of 35880 bp in length
102773 102872: gap of 100 bp
102873 106863: contig of 3991 bp in length
106864 106963: gap of 100 bp
106964 129804: contig of 22841 bp in length
129805 129904: gap of 100 bp
129905 132223: contig of 2319 bp in length
132224 133223: gap of 100 bp
133224 135764: contig of 3441 bp in length
135765 135864: gap of 100 bp
135865 143675: contig of 7811 bp in length.

FEATURES

Source
1. .143675
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone_lib="RPC1-11.1"
/clone="RP11-146P9"
misc_feature
1. .12163
/note="assembly_fragment:00440
vector_side:left"
clone_end:T7
12264..25269
misc_feature
/note="assembly_fragment:00434
fragment_chain:1"
25370..57709
/note="assembly_fragment:00526
fragment_chain:1"
57810..66792
misc_feature
/note="assembly_fragment:00661
fragment_chain:1"
66893..102772
misc_feature
/note="assembly_fragment:00323"
102873..106863
misc_feature
/note="assembly_fragment:00860"
106964..129804
misc_feature
/note="assembly_fragment:00978"
129905..132223
misc_feature
/note="assembly_fragment:01068"
132324..135764
misc_feature
/note="assembly_fragment:01424"
135865..143675
/note="assembly_fragment:01413
clone_end:SP6
vector_side:right"
BASE COUNT 42329 a 30042 c 28964 g 41431 t 909 others
ORIGIN

Query Match 5.4%; Score 100.2; DB 79; Length 143675;
Best Local Similarity 97.1%; Pred. No. 5.4e-14;
Matches 102; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 aqaagctgcacagaaacagagggagatttgcgtgcagccagagagaccag 60
|||||
Db 136065 AGNAGCTGCATCAGAAAAACAGAGGGGAGATTTGTGTGCTGCACCCGAGGAGACCAG 136124

QY 61 gaagatctcatggtgggaagacacctgatgatacagaggtgagaa 105
|||||
Db 136125 GAAGATCTCATGCTGGGAGGACCTGATGATACAGAGGTCTGTGA 136169

RESULT 8

AL358573
LOCUS 267581 bp DNA HTG 15-APR-2001
DEFINITION Homo sapiens chromosome 9 clone RP11-133022, *** SEQUENCING IN
PROGRESS ***, 37 unordered pieces.
ACCESSION AL358573
VERSION AL358573.17 GI:13660951
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Mclay, K.
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13398774.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bal33022
----- Summary Statistics
Assembly program: XCAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 256699 bases at least Q40
Consensus quality: 259744 bases at least Q30
Consensus quality: 261407 bases at least Q20
Insert size: 263981; sum-of-contigs
Insert size: 135491; 19.3% error; agarose-fp
Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality
coverage: 12.23x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 25718: contig of 25718 bp in length
25719 25818: gap of 100 bp
25819 29567: contig of 3749 bp in length
29568 29667: gap of 100 bp
29668 40520: contig of 10853 bp in length
40521 40620: gap of 100 bp
40621 50002: contig of 9382 bp in length
50003 50102: gap of 100 bp
50103 54942: contig of 4840 bp in length
54943 55042: gap of 100 bp
55043 59067: contig of 4025 bp in length
59068 59167: gap of 100 bp
59168 72327: contig of 13160 bp in length
72328 72427: gap of 100 bp
72428 79396: contig of 6969 bp in length
79397 79496: gap of 100 bp
79497 82738: contig of 3242 bp in length
82739 82838: gap of 100 bp
82839 87543: contig of 4705 bp in length
87544 87643: gap of 100 bp
87644 89847: contig of 2204 bp in length


```
* 89848 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7323 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length
* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2383 bp in length
* 121870 121969: gap of 100 bp
* 121970 125346: contig of 3577 bp in length
* 125347 125646: gap of 100 bp
* 125647 128990: contig of 3344 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length
* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2198 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.
```

FEATURES

source

```
1..267581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone_lib="RPC1-11.1"
1..25718
/note="assembly_fragment:00645"
fragment_chain:1
clone_end:77
vector_side:left"
25819..29567
/note="assembly_fragment:01660"
fragment_chain:1"
29668..40520
/note="assembly_fragment:03080"
fragment_chain:1"
40621..50002
/note="assembly_fragment:00771"
fragment_chain:1"
```

misc_feature

misc_feature

misc_feature

misc_feature

```
misc_feature
50103..54942
/note="assembly_fragment:02629"
fragment_chain:1"
55043..59067
/note="assembly_fragment:04565"
fragment_chain:1"
59168..72327
/note="assembly_fragment:00223"
fragment_chain:1"
72428..79396
/note="assembly_fragment:03318"
fragment_chain:1"
79497..82738
/note="assembly_fragment:00224"
fragment_chain:2"
82839..87543
/note="assembly_fragment:03050"
fragment_chain:2"
87644..89847
/note="assembly_fragment:01948"
fragment_chain:3"
89948..92376
/note="assembly_fragment:02191"
fragment_chain:3"
92477..99799
/note="assembly_fragment:00010"
99900..102117
/note="assembly_fragment:00672"
102218..109330
/note="assembly_fragment:01326"
109431..112187
/note="assembly_fragment:01498"
112288..116407
/note="assembly_fragment:01607"
116508..119386
/note="assembly_fragment:01899"
119487..121869
/note="assembly_fragment:02068"
121970..125546
/note="assembly_fragment:02469"
125647..128990
/note="assembly_fragment:02502"
129091..134055
/note="assembly_fragment:02669"
134156..138314
/note="assembly_fragment:02757"
138415..140612
/note="assembly_fragment:02776"
140713..143296
/note="assembly_fragment:03127"
143397..148113
/note="assembly_fragment:03634"
148214..151009
/note="assembly_fragment:03898"
151110..156616
/note="assembly_fragment:04220"
156717..158763
/note="assembly_fragment:04276"
158864..162129
/note="assembly_fragment:04418"
162230..164443
/note="assembly_fragment:04428"
```

Query Match

5.4%; Score 100.2; DB 80; Length 267581;

Best Local Similarity 97.1%; Pred. No. 5.8e-14;

Matches 102; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 aqaagctgcatcagaaacagagggagatttggctgcagccagggagaccag 60

|||||

Db 253275 AGNAGCTGCATCAGAAAAACAGAGGGGAGATTGTTGGCTGCACCCGAGGAGACCAG 253334

|||||

Qy 61 gaagatctgcatggtgggaaggacctgatcatagaggtgagaa 105

|||||

Db 253335 GAAGATCTGATGGTGGGAAGGACCTGATGATACAGAGGTCTGTA 253379

RESULT 9
AC009556/c

LOCUS
DEFINITION

AC009556 172298 bp DNA HTG 26-MAY-2000
Homo sapiens clone RP11-57C21, WORKING DRAFT SEQUENCE, 26 unordered
pieces.

AC009556

AC009556.4

GI:7107752

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

2137 2236: gap of 100 bp
2237 3571: contig of 1335 bp in length
3572 3671: gap of 100 bp
3672 5272: contig of 1601 bp in length
5273 5372: gap of 100 bp
5373 6381: contig of 1009 bp in length
6382 6481: gap of 100 bp
6482 8508: contig of 2027 bp in length
8509 8608: gap of 100 bp
8609 9886: contig of 1278 bp in length
9887 9986: gap of 100 bp
9987 11943: contig of 1957 bp in length
11944 12043: gap of 100 bp
12044 15112: contig of 3069 bp in length
15113 15212: gap of 100 bp
15213 19323: contig of 4111 bp in length
19324 19423: gap of 100 bp
19424 24065: contig of 4642 bp in length
24066 24165: gap of 100 bp
24166 31658: contig of 7493 bp in length
31659 31758: gap of 100 bp
31759 38737: contig of 6979 bp in length
38738 38837: gap of 100 bp
38838 46131: contig of 7294 bp in length
46132 46231: gap of 100 bp
46232 52344: contig of 6113 bp in length
52345 52444: gap of 100 bp
52445 59763: contig of 7319 bp in length
59764 59863: gap of 100 bp
59864 66816: contig of 6953 bp in length
66817 66916: gap of 100 bp
66917 74010: contig of 7094 bp in length
74011 74110: gap of 100 bp
74111 80599: contig of 6489 bp in length
80600 80699: gap of 100 bp
80700 89507: contig of 8808 bp in length
89508 89607: gap of 100 bp
89608 100337: contig of 10730 bp in length
100338 100437: gap of 100 bp
100438 114298: contig of 13861 bp in length
114299 114398: gap of 100 bp
114399 128530: contig of 14132 bp in length
128531 128630: gap of 100 bp
142918 143017: contig of 14287 bp in length
143018 156243: contig of 13226 bp in length
156244 156343: gap of 100 bp
156344 172298: contig of 15955 bp in length.

FEATURES

source

1..172298
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-57C21"
/clone_lib="RPC1-11 Human Male BAC"
1..1016
/note="assembly_fragment"
1117..2136
/note="assembly_fragment"
2237..3571
/note="assembly_fragment"
3672..5272
/note="assembly_fragment"
5373..6381
/note="assembly_fragment"
6482..8508
/note="assembly_fragment"
8609..9886
/note="assembly_fragment"
9987..11943
/note="assembly_fragment"
12044..15112
/note="assembly_fragment"
15213..19323

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

TITLE

JOURNAL

COMMENT

Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced gi:6479158.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1672

Center clone name: 57_C_21

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-primer-amersham; 5% of reads

Chemistry: Dye-terminator Big Dye; 95% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136385 bases at least Q40

Consensus quality: 156543 bases at least Q30

Consensus quality: 165534 bases at least Q20

Insert size: 165000; agarose-fp

Insert size: 169798; sum-of-contigs

Quality coverage: 4.2 in Q20 bases; agarose-fp

Quality coverage: 4.1 in Q20 base.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 26 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1016: contig of 1016 bp in length

* 1017 1116: gap of 100 bp

* 1117 2136: contig of 1020 bp in length

misc_feature /note="assembly_fragment"
19424..24065
misc_feature /note="assembly_fragment"
24166..31658
misc_feature /note="assembly_fragment"
31759..38737
misc_feature /note="assembly_fragment"
38838..46131
misc_feature /note="assembly_fragment"
46232..52344
misc_feature /note="assembly_fragment"
52445..59763
misc_feature /note="assembly_fragment"
59864..66816
misc_feature /note="assembly_fragment"
66917..74010
misc_feature /note="assembly_fragment"
vector_side:left
vector_side:right
74111..80599
misc_feature /note="assembly_fragment"
clone_end:SP6
vector_side:left
80700..89507
misc_feature /note="assembly_fragment"
89608..100337
misc_feature /note="assembly_fragment"
100438..114298
misc_feature /note="assembly_fragment"
114399..128530
misc_feature /note="assembly_fragment"
128631..142917
misc_feature /note="assembly_fragment"
143018..156243
misc_feature /note="assembly_fragment"
156344..172298
misc_feature /note="assembly_fragment"
BASE COUNT 51330 a 34914 c 34487 g 49065 t 2502 others
ORIGIN

Query Match 5.3%; Score 98.6; DB 61; Length 172298;
Best Local Similarity 96.2%; Pred. No. 1.4e-13;
Matches 101; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 agaagctggcatcagaaaaacagagggagatttggctgcagcgaggagagaccag 60
Db 16539 AGAAGCTGGCATCAGAAAAACAGAGGGGAGATTGTGTGGCTGCAGCGAGGAGACCAG 16480
Qy 61 gaagatctgcattgggaagacacctgatgatacagagtgagaa 105
Db 16479 GAAGATTGCATGGTGGGAGGACCTGATGATACAGAGTCTGTA 16435

RESULT 10
AF279290
LOCUS AF279290 580 bp DNA PRI 04-DEC-2000
DEFINITION Homo sapiens prostate-cancer-specific DD3 protein gene, promoter region and exon 1.
ACCESSION AF279290
VERSION AF279290.1 GI:11528086
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 580)
AUTHORS Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and Bussemakers,M.J.G.
TITLE Isolation and Characterization of the Promoter of the Human Prostate Cancer-specific DD3 Gene
J. Biol. Chem. 275 (48), 37496-37503 (2000)
PUBMED 10982808

REFERENCE 2 (bases 1 to 580)
AUTHORS Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and Bussemakers,M.J.G.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-2000) Urology Research Laboratory, University Medical Center Nijmegen, Geert Grooteplein Zuid 10, Nijmegen 6525 GA, The Netherlands
FEATURES Location/Qualifiers
source 1..580
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q21-q22"
1..460
promoter 260..288
protein_bind /bound_moiety="high mobility group I Y protein"
461..580
mRNA /product="prostate-cancer-specific DD3 protein"
461..580
exon /number=1
BASE COUNT 190 a 97 c 153 g 140 t
ORIGIN
Query Match 5.2%; Score 98; DB 89; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 agaagctggcatcagaaaaacagagggagatttggctgcagcgaggagagaccag 60
Db 483 AGAAGCTGGCATCAGAAAAACAGAGGGGAGATTGTGTGGCTGCAGCGAGGAGACCAG 542
Qy 61 gaagatctgcattgggaagacacctgatgatacagag 98
Db 543 GAAGATCTGCATGGTGGGAGGACCTGATGATACAGAG 580
RESULT 11
I66494
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
source 1..7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match 3.2%; Score 59.8; DB 10; Length 7218;
Best Local Similarity 5.1%; Pred. NO. 0.00039;
Matches 20; Conservative 219; Mismatches 155; Indels 0; Gaps 0;
Qy 1013 ttcatctccacttttgccattctcaagacacctcaaatgtcattccattaatatcac 1072
Db 1076 YY 1135
Qy 1073 aggattaaactttttttaacctggaagaattcaatgttacatgcagctatgggaatt 1132
Db 1136 YY 1195
Qy 1133 aattacatatattgtttccagtcgaaagatgactaaagtccttatccctccctttgt 1192
Db 1196 YY 1255

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 04:11:46 ; Search time 6114.61 Seconds
(without alignments)
1267.675 Million cell updates/sec

Title: US-09-402-713A-4
Perfect score: 820
Sequence: 1 agaagctggcatcagaaaaa.....cattactcattgttcaaa 820

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estrol:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estrol0:*
92: em_estrol1:*
93: em_estrol2:*
94: em_estrol3:*
95: em_estrol4:*
96: em_estrol5:*
97: em_estrol6:*
98: em_estrol7:*
99: em_estrol8:*
100: em_estrol9:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.


```

DEFINITION RC5-FT0193-201100-012-D06 FT0193 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF858286
VERSION BF858286.1 GI:12246030
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 332)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-FT0193-
201100-012-D06&t3=2000-11-20&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 332.
Location/Qualifiers
1. 332
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0193"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 90 a 84 c 64 g 94 t
ORIGIN
Query Match 35.5%; Score 290.8; DB 170; Length 332;
Best Local Similarity 99.3%; Pred. No. 3.2e-71;
Matches 292; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 527 tgcacaaacaaatgaatactgtgatgacatgagcgagcgcaagctggggaggagataa 586
|||||
Db 12 TGCACAAACAAATGAATACTGTGATGACATGAGCGAGCGCAAGCTGGGAGGAGATAA 71
|||||

QY 587 ccacgggagagggctcaggattctgctgctgcctaaactgtcgttcataaccaa 646
|||||
Db 72 CCACGGGGCAGAGGGTCAGAGATTCTGGCCCTGCTGCCTAACTGTGCGTTCAATACCAA 131
|||||

QY 647 tcaattcatttttaacacctcaaaacaaagctgttgtaatatctgtactctacgltcc 706
|||||
Db 132 TCATTTTCATATTTCTAACCTCAAAACAAAGCTGTGTGTAATATCTGATCTCTAGGGTTCC 191
|||||

QY 707 ttctggggccaacattctccatatatccagccacactcatttttaattagttccag 766
|||||
Db 192 TTCGGGGCCAAACATTTCTCCATATATCCAGCCACACTCATTTTAAATATTAGTTCACAG 251
|||||

QY 767 atctgtactgtgacctttctacactgtagaataaacattactctatttcttcaaa 820
|||||
Db 252 ANCTGTACTGTGACCTTCTACGCTGTAGTAACAACTACTCATTTTGTTCAAA 305
|||||

```

RESULT 3

```

BF373619 290 bp mRNA EST 24-NOV-2000
MR0-FT0175-310800-106-h09 FT0175 Homo sapiens cDNA, mRNA sequence.
BF373619
VERSION BF373619.1 GI:11335644
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 290)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-FT0175-
310800-106-h09&t3=2000-08-31&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 290.
Location/Qualifiers
1. 290
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0175"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 78 a 73 c 53 g 86 t
ORIGIN
Query Match 29.8%; Score 244.4; DB 147; Length 290;
Best Local Similarity 99.6%; Pred. No. 3.7e-58;
Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 575 gggagagataaccacggggcgaggggtcaggattctgcccctgctgcctaactgtcg 634
|||||
Db 21 GGGAGGAGATAACACGGGGCAGAGGGTCAGGATTCTGGCCCTGCTGCCTAACTGTGGC 80
|||||

QY 635 ttcataaccaaattcatttcatttcctaacctcaaaacaaagctgttgtaatatctgat 694
|||||
Db 81 TTCATAACCAATTCATTCATATTTCTAACCTCAAAACAAAGCTGTGTGTAATATCTGAT 140
|||||

QY 695 ctctacgggttcctctggggcccaacattctccatatatccagccacactcttttaata 754
|||||
Db 141 CTCTACGGTTCCTTCTGGGCCCAACATTTCTCATATATCCAGCCACACTCATTTTAAATA 200
|||||

QY 755 tttagttccagatctgtactgtgacctttctacactgtagaataaacattactctatttg 814
|||||

```



```

QY 682 tgaataatgatctctacaggttcccttctcttggtggcccaacattccatataatccagcaca 741
|||||
Db 166 TGAATAATCTGATCTCTACGGTTCCTTCTGGGCCCAACATCTCCATATATCCAGCCACA 107
|||||
QY 742 ctcatTTtaatttagttccagatctgtactgtgacctttctacactgtagaataac 801
|||||
Db 106 CTCAATTTTAATATTTAGTTCCAGATCTGTACTGTGACCTTCTTACACTGTAGATAAC 47
|||||
QY 802 attactcattttgttcaaa 820
|||||
Db 46 ATTACTCATTTTGTGTCAAA 28
|||||

RESULT 6
BF373406 167 bp mRNA EST 24-NOV-2000
DEFINITION IL2-FT0159-070800-120-H01 FT0159 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF373406
VERSION BF373406.1 GI:11335431
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 167)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL2&t2=IL2-FT0159-
070800-120-H01&t3=2000-08-07&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 167.
FEATURES
source
1..167
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0159"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site:1: SmaI;
Site:2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 47 a 38 c 49 g 33 t
ORIGIN

Query Match 20.4%; Score 167; DB 147; Length 167;
Best Local Similarity 100.0%; Pred. No. 2.2e-36;
Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 cacagaagcacaaaaggacacagatccctggagaaatgcccgccgacctgtg 313
|||||

```

```

Db 1 CACAGGACGACAAAAGGAGGACAGAGATCCCTGGAGAAATGCCGGCGCCATCTTG 60
QY 314 ggtcatcgatgagcctgcctgtgctccgtgtccgcgttgagggaaggacattagaaaa 373
|||||
Db 61 GGTCACTGATGAGCCTCGCCCTGTGCTGTGCTCCGCTTGTGAGGAAGACATTAGAAA 120
|||||
QY 374 tgaatcgatgttctcttaagagatggcaggagaaacagatcctgtt 420
|||||
Db 121 TGAATTGATGTGTCTCTTAAGGATGGCGAGGAAAACAGATCCTGTT 167
|||||

RESULT 7
AQ206972/2 394 bp DNA GSS 17-SEP-1998
LOCUS HS_3238_B1_G11_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3238 Col=21 Row=N, DNA sequence.
ACCESSION AQ206972
VERSION AQ206972.1 GI:3617542
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 394)
Mahairs,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairs GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3238 row: N column: 21
Class: BAC ends
High quality sequence stop: 394.
FEATURES
source
1..394
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=3238 Col=21 Row=N"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 134 a 62 c 93 g 103 t 2 others
ORIGIN

Query Match 13.7%; Score 112.4; DB 225; Length 394;
Best Local Similarity 97.4%; Pred. No. 7.1e-21;
Matches 113; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 705 cctctggcccaacatttcacatataatcagccacacatttttaattagttccc 764
|||||
Db 394 CCGCTGGGCNCACATCTCCATATATCCAGCCACACTATTTTAATATTAGTTCCC 335
|||||
QY 765 agatctgtactgtgaccttctcacatgtagaataacattactctattgttcaaa 820
|||||
Db 334 AGATCTGTACTGTGACCTTCTTACACTGTAGATAACATTACTCATTTTGTTCAAA 279
|||||

RESULT 8
AI557495 657 bp mRNA EST 09-AUG-1999
LOCUS PT2.1.7.H12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
DEFINITION AI557495
ACCESSION AI557495

```



```

/db_xref="taxon:10090"
/clone="IMAGE:355452"
/clone_lib="Soares mouse embryo NDBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGGAGCGCGGAAATTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 l; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Fatima Bonaldo."
BASE COUNT      122 a 131 c 119 g 109 t
ORIGIN

Query Match      4.9%; Score 40.4; DB 20; Length 481;
Best Local Similarity 61.3%; Pred. No. 1.5;
Matches 65; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 447 agatttgaatgaagtccaaagtgcacattaccacgaatgagaggaacacagcgagaaat 506
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AATTGTAGGAAGAAGAAAAAAGAAATTTAAAAAAGAAAAAAGAAAAAAGAAAGATGAC 60

QY 507 ctgtggtgcttcacagacatgcacacacaaatggaaatctgtg 552
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 CTGTATGGAATAATTTAAAAAAGAAAAAAGAAAGATATATCTGTG 106

Search completed: July 31, 2001, 04:11:52
Job time: 10004 sec

```

RECORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

house mouse.
 Mus musculus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 481)
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Underwood, K., Steptoe, M., Thelasing, B., Allen, M., Bowers, Y., Person
 , B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter
 , E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterston, R. and Wilson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouse@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)
 MGI:227252
 Seq primer: -40RP from Gibco
 High quality sequence stop: 454
 POLYA=No. Location/Qualifiers
 1. .481
 /organism="Mus musculus"
 /strain="C57BL/6J"

FEATURES
source

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:20:07 ; Search time 176.8 seconds
(without alignments)
878.027 Million cell updates/sec

Title: US-09-402-713A-4
Perfect score: 820
Sequence: 1 agaagctgcacagaaaaa.....cattactcattgttcaaa 820

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	57.4	7.0	7218	1	US-08-232-463-14
2	35.6	4.3	7218	1	US-08-232-463-14
3	34.2	4.2	2679	1	US-07-977-434-11
4	34.2	4.2	2679	1	US-08-458-819-11
5	34.2	4.2	2679	5	PCT-US91-07035-11
C 6	33.8	4.1	246240	2	US-08-724-394A-20
C 7	33.8	4.1	246240	2	US-08-724-394A-21
C 8	33.8	4.1	246240	2	US-08-724-394A-22
C 9	33.4	4.0	713	3	US-08-532-896-26
10	31	3.8	2737	3	US-08-389-564B-5
11	31	3.8	2737	3	US-08-466-047B-5
12	30.6	3.7	466	3	US-08-899-437-21
13	30.6	3.7	466	4	US-09-126-121-21
14	30.6	3.7	2091	3	US-08-899-437-22
15	30.6	3.7	2091	4	US-09-126-121-22
16	30.6	3.7	2502	3	US-08-899-437-5
17	30.6	3.7	2502	4	US-09-126-121-5
18	30.6	3.7	4758	3	US-09-191-647-1
C 19	30.6	3.7	43795	3	US-08-743-185-101
20	30.4	3.7	877	3	US-09-128-888-1
21	30.4	3.7	2233	1	US-08-129-129-2
C 22	30.2	3.7	29604	3	US-08-781-891-207
C 23	29.8	3.6	9997	1	US-08-246-982A-15
C 24	29.8	3.6	9997	1	US-08-453-265-15
C 25	29.8	3.6	10103	2	US-08-457-273B-7
C 26	29.8	3.6	10607	1	US-08-078-090-3
C 27	29.8	3.6	35100	1	US-08-306-691B-19

C 28	29.8	3.6	35100	5	PCT-US93-06251-19	Sequence 19, Appl
C 29	29.6	3.6	2259	1	US-07-828-700-7	Sequence 7, Appl
30	29.4	3.6	320	1	US-08-171-385-18	Sequence 18, Appl
31	29.4	3.6	320	3	US-08-361-441B-18	Sequence 18, Appl
C 32	29.4	3.6	879	4	US-09-023-213B-1	Sequence 1, Appl
C 33	29.4	3.6	917	4	US-08-942-012B-1	Sequence 1, Appl
C 34	29.4	3.6	924	1	US-07-593-657-4	Sequence 4, Appl
35	29.4	3.6	1834	1	US-08-297-633A-1	Sequence 1, Appl
36	29.4	3.6	1834	1	US-08-485-721-8	Sequence 8, Appl
37	29.4	3.6	1834	2	US-08-392-935-8	Sequence 8, Appl
38	29.4	3.6	1834	5	PCT-US93-08325-1	Sequence 1, Appl
39	29.4	3.6	1834	5	PCT-US93-08326-8	Sequence 8, Appl
C 40	29.4	3.6	5430	3	US-09-012-515A-11	Sequence 11, Appl
C 41	29.4	3.6	5430	3	US-08-360-144A-11	Sequence 11, Appl
C 42	29.4	3.6	6755	3	US-08-931-999-4	Sequence 4, Appl
C 43	29.4	3.6	7824	5	PCT-US95-06722-11	Sequence 11, Appl
C 44	29.2	3.6	455	2	US-08-680-395-8	Sequence 8, Appl
C 45	29.2	3.6	1162	1	US-08-474-140-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 INMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base-pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT99pt-F1s
US-08-232-463-14


```

; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cserr
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: Case No. 5466591 8753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2972
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2679 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Thermosipho africanus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2676
; US-07-977-434-11

Query Match 4.2%; Score 34.2; DB 1; Length 2679;
Best Local Similarity 54.3%; Pred. No. 0.75;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 102 agaaataagaagggctgctgactttaccatctgagggccacacatctgctgaaatggagat 161
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2415 AGGAGAAAGAAATAGCTGTTAAACACTCCATTCACGAAACAGCAGCTCATATAAAGAT 2474
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 162 aattaacatcactagaacagcaagatgacaataataatgctctaagttagtgcacatgtttt 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2475 AGCTATGATTATATTCATAATAGATTGAAGAAGGAAAATCTACGTTCAAAAATGATATT 2534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 222 gcacatt 228
    ||| ||
DB 2535 GCAGGTT 2541

RESULT 4
US-08-458-819-11
; Sequence 11, Application US/08458819
; Patent No. 5795762
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abranson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; ZIP: 07110-1199
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7
; SOFTWARE: WordPerfect 2.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,819
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/977,434
; FILING DATE: 23-FEB-1993
; APPLICATION NUMBER: US 590,490
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,466
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,213
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 523,394
; FILING DATE: 15-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 143,441
; FILING DATE: 12-JAN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 063,509
; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cserr
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: Case No. 5795762 8753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2972
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2679 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Thermosipho africanus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2676
; US-07-977-434-11
```


Db 36325 TTAAGTGAAGTCTCTGAA 36309

RESULT 9
US-08-532-896-26/c
: Sequence 26, Application US/08532896
: Patent No. 6124115
: GENERAL INFORMATION:
: APPLICANT: LUU-THE, Van
: APPLICANT: LABRIE, Fernand
: TITLE OF INVENTION: PRODUCTION AND USE OF ISOLATED TYPE 2
: TITLE OF INVENTION: 17B-HYDROXYSTEROID DEHYDROGENASE
: NUMBER OF SEQUENCES: 59
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
: STREET: 1180 Avenue of the Americas
: CITY: New York
: STATE: NY
: COUNTRY: US
: ZIP: 10036-8403
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/532, 896
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Meilman, Edward
: REGISTRATION NUMBER: 24,735
: REFERENCE/DOCKET NUMBER: P/1259-313
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 382-0700
: TELEFAX: (212) 382-0888
: TELEX: 236925
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 713 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: PS-08-532-896-26

		Query Match	4.0%; Score 32.4; DB 3; Length 713;	
		Best Local Similarity	48.4%; Pred. No. 1.3;	
		Matches	90; Conservative 0; Mismatches 96; Indels 0; Gaps 0;	
Qy	93	acagagtgaataagaaggctgctgacctttaccatctgagggccacacatctgtcga	152	
Dd	501	AAAAACATTTGAAATAAATAAGATGCTGAATTGCAGGATCTAGCAACAATTCCATAT	442	
Qy	153	aatggagataattaacatcatcagtaaacacagcaaatgatacaatatattgtctaagttagtga	212	
Dd	441	GAAGGAGGTGGGAATTGGCACCTCGACCACAAAAGTGTGATGGAATGTTTTTGATTGTGATA	382	
Qy	213	catgtttttgcacatttcacgcccccctttaatatccacacacacaggaagcacaaaaagga	272	
Dd	381	AGGAGCTGAGCAGGTTTGGGACAATGTGATATAAGCACAGATTTGGAAACATCTAATGAA	322	
Qy	273	agcata 278 		
Dd	321	GTCACA 316		
RESULT 10				
US-08 - 389 - 564B - 5				
Sequence 5. Application US/08389564B				

APPLICANT: Nevalainen, Kaisu Milja Helena
 APPLICANT: Penttil, Merja E.
 TITLE OF INVENTION: Fungal Promoters Active In The Presence
 OF Glucose
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 STREET: 1100 New York Avenue, Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/466,047B
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/389,564
 FILING DATE: 16-FEB-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/932,564
 FILING DATE: 19-AUG-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/496,155
 FILING DATE: 19-MAR-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/044,077
 FILING DATE: 29-APR-1987
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 86 10600
 FILING DATE: 30-APR-1986
 ATTORNEY/AGENT INFORMATION:
 NAME: REED, GRANT E.
 REGISTRATION NUMBER: 41,264
 REFERENCE/DOCKET NUMBER: 1716.008000H
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2737 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-466-047B-5

```

Query Match      3.8%;   Score 31;   DB 3;   Length 2737;
Best Local Similarity 56.3%;   Pred. No. 7.6;
Matches 58;   Conservative 0;   Mismatches 45;   Indels 0;   Gaps 0;
Qy  50  agggagaccaggaagatctgcattgtgggaaggacctgatcatcacagaggtgagaataaa 109
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  2056  AGAGGCATCTGGAGGACAAGTATGGGGGGAGGGATGACCATTAAGTTCGGGGTGATTA 2115
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy  110  gaaaggctctgactttaccatctgagggccacacatctgctga 152
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  2116  CCATGACTCGCATCGGACGAAGTGAGGGGGGCATCTGCTCA 2158
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-08-899-437-21
; Sequence 21, Application US/08899437
; Patent No. 6121415
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor

```

```

; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,437
; FILING DATE: 24-Jul-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 466 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: EST Genbank entry H23651
; LOCATION: 1-466
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; US-09-126-121-21

Query Match 3.7%; Score 30.6; DB 3; Length 466;
Best Local Similarity 53.8%; Pred. No. 3.8;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 450 ttgaaatgaagtcacaaagtgcattaccatgagagaaacacagacgagaaaatctt 509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TTCAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 510 gatgcttcacagacatcaacaaatgaatactgtgatgacatgagcgag 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 AAAAGCTACAGTCTCAAGCATCCAGCAACAAATGCGAAAGTCAGAGAACTTGGTGAAG 275

RESULT 13
US-09-126-121-21
; Sequence 21, Application US/09126121
; Patent No. 6252051
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/09/126,121
; FILING DATE: 30-Jul-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 466 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: EST Genbank entry H23651
; LOCATION: 1-466
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; US-09-126-121-21

Query Match 3.7%; Score 30.6; DB 4; Length 466;
Best Local Similarity 53.8%; Pred. No. 3.8;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 450 ttgaaatgaagtcacaaagtgcattaccatgagagaaacacagacgagaaaatctt 509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TTCAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 510 gatgcttcacagacatcaacaaatgaatactgtgatgacatgagcgag 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 AAAAGCTACAGTCTCAAGCATCCAGCAACAAATGCGAAAGTCAGAGAACTTGGTGAAG 275

RESULT 14
US-08-899-437-22
; Sequence 22, Application US/08899437
; Patent No. 6121415
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,437
; FILING DATE: 24-Jul-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2091 base pairs
; TYPE: Nucleic Acid
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:26:09 ; Search time 478.87 Seconds
(without alignments)
1075.196 Million cell updates/sec

Title: US-09-402-713A-4
Perfect score: 820
Sequence: 1 agagctgcatcagaaaa.....cattactcattttgttcaaa 820

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues 1460202
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	820	100.0	820	19 AAV62429	Prostate cancer an
2	820	100.0	1872	19 AAV62428	Prostate cancer an
3	812	99.0	812	21 AAA06690	Human immunogenic
4	725.8	88.5	3112	21 AAA06687	Human immunogenic
5	725.6	88.5	2229	21 AAA06688	Human immunogenic
6	725.6	88.5	2426	21 AAA06689	Human immunogenic
7	725.4	88.5	2037	19 AAV62427	Prostate cancer an
8	725.4	88.5	3582	19 AAV62430	Prostate cancer an
9	523.6	63.9	597	20 AAX37486	Human secreted pro
10	513.2	62.6	718	21 AAA06545	Human immunogenic
11	414.2	50.5	437	21 AAC05768	Human secreted pro

c 12	330	40.2	359	20	AAZ33445	Human prostate can
c 13	288.4	35.2	301	21	AAA06520	Human immunogenic
c 14	100	12.2	123	19	AAV33791	Prostate cancer an
c 15	76.4	9.3	936	22	AAF58252	Oligonucleotide D1
c 16	76.4	9.3	936	22	AAF58254	Oligonucleotide D1
c 17	76.4	9.3	936	22	AAF58257	Oligonucleotide D1
c 18	76.4	9.3	936	22	AAF58259	Oligonucleotide D2
c 19	76.4	9.3	936	22	AAF58262	Oligonucleotide D2
c 20	76.4	9.3	938	22	AAF58255	Oligonucleotide D1
c 21	71	8.7	936	22	AAF58252	Oligonucleotide D1
c 22	71	8.7	936	22	AAF58254	Oligonucleotide D1
c 23	71	8.7	936	22	AAF58257	Oligonucleotide D1
c 24	71	8.7	936	22	AAF58259	Oligonucleotide D2
c 25	71	8.7	936	22	AAF58262	Oligonucleotide D2
c 26	71	8.7	938	22	AAF58255	Oligonucleotide D1
c 27	39.8	4.9	244	22	AAF58238	Oligonucleotide D1
c 28	38.4	4.7	244	22	AAF58238	Oligonucleotide D1
c 29	36	4.4	254	21	AAC16115	Human secreted pro
c 30	35.8	4.4	67212	21	AAA08954	WFS1 variant genom
c 31	35.2	4.3	2147	21	AAZ60617	DNA encoding the p
c 32	34.6	4.2	982	20	AAV37417	Human secreted pro
c 33	34.2	4.2	1830	13	AAQ24334	Mutant thermostabl
c 34	34.2	4.2	2073	13	AAQ24333	Mutant thermostabl
c 35	34.2	4.2	2265	13	AAQ24332	Mutant thermostabl
c 36	34.2	4.2	2403	13	AAQ24331	Mutant thermostabl
c 37	34.2	4.2	2568	13	AAQ28937	Encodes Taf DNA po
c 38	34.2	4.2	2571	13	AAQ24330	Mutant thermostabl
c 39	34.2	4.2	2679	13	AAQ24329	Mutant thermostabl
c 40	34.2	4.2	2679	13	AAQ28936	Encodes Asp37 Taf
c 41	34.2	4.2	4286	13	AAQ23917	Taf DNA polymerase
c 42	34.2	4.2	122186	22	AAC89560	Human histone deac
c 43	33.8	4.1	261	21	AAC10644	Human secreted pro
c 44	33.8	4.1	1257	21	AAC74385	Human secreted pro
c 45	33.8	4.1	235033	19	AAV57926	Hereditary haemoch

ALIGNMENTS

RESULT 1
AAV62429
ID AAV62429 standard; cDNA; 820 BP.
XX
AC AAV62429;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 3.
XX
KW Prostate cancer antigen cDNA splice variant 3; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
(DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
XX prevention and treatment of prostatic cancer
XX Claim 4; Pages 77-78; 111pp; English.

CC The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 3 sequence comprising of exons 1, 3, and 4a
CC of the PCA3 gene. The PCA3 cDNA splice variant 3 sequence,
CC isolated from a human primary prostatic tumour tissue cDNA library,
CC was found in approximately 15% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunosay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (ant)agonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 820 BP; 262 A; 169 C; 191 G; 198 T; 0 other;

Query Match 100.0%; Score 820; DB 19; Length 820;
Best Local Similarity 100.0%; Pred. No. 3.3e-249;
Matches 820; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaagctggcatcagaaaaacagaggaggagatttggctcagccgaggagaccag 60
DB 1 agaagctggcatcagaaaaacagaggaggagatttggctcagccgaggagaccag 60
QY 61 gaagatctgatgtgggaaggacccctgatgatacagaggtgagaaataagaaggtcgt 120
DB 61 gaagatctgatgtgggaaggacccctgatgatacagaggtgagaaataagaaggtcgt 120
QY 121 gactttaccatctgagggccacacatctgctgaaatggagataattaacatcactagaac 180
DB 121 gactttaccatctgagggccacacatctgctgaaatggagataattaacatcactagaac 180
QY 181 agcaatgacacaataatgtctaaagttagtgacatgtttttgacatttccagcccttt 240
DB 181 agcaatgacacaataatgtctaaagttagtgacatgtttttgacatttccagcccttt 240
QY 241 aaatattccacacacaggaagcacaaaaaggaagcagagatccctgggagaaatgcc 300
DB 241 aaatattccacacacaggaagcacaaaaaggaagcagagatccctgggagaaatgcc 300
QY 301 ggcgcgcatttgggttcacgatgagcctgcctctgctgctgctgctgctgctgctgct 360
DB 301 ggcgcgcatttgggttcacgatgagcctgcctctgctgctgctgctgctgctgctgct 360
QY 361 ggcattagaaatgaatgatgtgttccttaagagatggcaggagaaacagatctgtt 420
DB 361 ggcattagaaatgaatgatgtgttccttaagagatggcaggagaaacagatctgtt 420
QY 421 gtgattatttattgacgggattacagatttgaatgaagtacacaaagtgcattacc 480
DB 421 gtgattatttattgacgggattacagatttgaatgaagtacacaaagtgcattacc 480
QY 481 aatgagagaaacacagacagaaaaatttgatggcttcacaaagacatgcacaaacaaa 540
DB 481 aatgagagaaacacagacagaaaaatttgatggcttcacaaagacatgcacaaacaaa 540
QY 541 tggaaactgtgatgacatgagcagcccaagctgggagagagataaccacggggcagag 600
DB 541 tggaaactgtgatgacatgagcagcccaagctgggagagagataaccacggggcagag 600
QY 601 gtcagattctgcccctgtcgttaactgtgcgttcataaaccacaaatcatttcatttc 660
DB 601 gtcagattctgcccctgtcgttaactgtgcgttcataaaccacaaatcatttcatttc 660
QY 661 taacctctaaacaaagctgtgttaatatctgatctctcaggttctcttcttgggcccaaca 720

DB 661 taacctctaaacaaagctgtgttaatatctgatctctcaggttctcttcttgggcccaaca 720
QY 721 ttctccatatatccagccacacacacatttttaatttagttccagatctgtactgtgac 780
DB 721 ttctccatatatccagccacacacacatttttaatttagttccagatctgtactgtgac 780
QY 781 cttctacactgtagaataacattactcattttgttcaaa 820
DB 781 cttctacactgtagaataacattactcattttgttcaaa 820
RESULT 2
AAV62428
ID AAV62428 standard; cDNA; 1872 BP.
XX
AC AAV62428;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 2.
XX
KW Prostate cancer antigen cDNA splice variant 2; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
XX prevention and treatment of prostatic cancer
XX
XX Claim 4: Pages 76-77; 111pp; English.
XX
CC The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 2 sequence comprising of exons 1, 3, 4a and
CC 4b of the PCA3 gene. The PCA3 cDNA splice variant 2 sequence,
CC isolated from a human primary prostatic tumour tissue cDNA library,
CC was found in approximately 65% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunosay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (ant)agonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 1872 BP; 567 A; 389 C; 369 G; 539 T; 8 other;

Query Match 100.0%; Score 820; DB 19; Length 1872;
Best Local Similarity 100.0%; Pred. No. 5.3e-249;
Matches 820; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 aagaagctgcatcagaaaaacagagggagatttgtgtgctgcagccagggagaccag 60
DB 1 aagaagctgcatcagaaaaacagagggagatttgtgtgctgcagccagggagaccag 60
QY 61 gaagatctgcattggtgggaagacctgatgatcacagaggtgagaaataagaaggctgt 120
DB 61 gaagatctgcattggtgggaagacctgatgatcacagaggtgagaaataagaaggctgt 120
QY 121 gactttaccatctgagggcacacactctgctgaaatggagataatlaacatcactagaaac 180
DB 121 gactttaccatctgagggcacacactctgctgaaatggagataatlaacatcactagaaac 180
QY 181 agcaagatgacaataatgtctaagtagtgacatgtttttgcacatttccagcccttt 240
DB 181 agcaagatgacaataatgtctaagtagtgacatgtttttgcacatttccagcccttt 240
QY 241 aaatatccacacacagagaagcaaaaggaagcacagagatccctgggagaaatgcc 300
DB 241 aaatatccacacacagagaagcaaaaggaagcacagagatccctgggagaaatgcc 300
QY 301 gcccgcattctgggtcatcagatgagcctgcctgtgcctgttcccgcttggaggaa 360
DB 301 gcccgcattctgggtcatcagatgagcctgcctgtgcctgttcccgcttggaggaa 360
QY 361 ggacattagaaaatgaattgattgttccctaaaggatggcagagaaacagatcctgt 420
DB 361 ggacattagaaaatgaattgattgttccctaaaggatggcagagaaacagatcctgt 420
QY 421 gtggatatattttgaacgggattacagatttgaatgaattgaattgaattgaattac 480
DB 421 gtggatatattttgaacgggattacagatttgaatgaattgaattgaattgaattac 480
QY 481 aatgagagaaaacagagagaaaatcttgatggtttcacaagacatgcaacaacaa 540
DB 481 aatgagagaaaacagagagaaaatcttgatggtttcacaagacatgcaacaacaa 540
QY 541 tggaaactgtgatgacatgagcgagcgaactgtgggagagataaacacggggcagag 600
DB 541 tggaaactgtgatgacatgagcgagcgaactgtgggagagataaacacggggcagag 600
QY 601 gtcaggattctggcctgcctgctaaactgtgcttataacaaatcattcatttcc 660
DB 601 gtcaggattctggcctgcctgctaaactgtgcttataacaaatcattcatttcc 660
QY 661 taacctcaaaaacaaagctgttgaatatctgtacgttctctctgttctgtggccaa 720
DB 661 taacctcaaaaacaaagctgttgaatatctgtacgttctctctgttctgtggccaa 720
QY 721 ttctccatatatccagccacactcatttttaattatttagttccagatctgtactgtgac 780
DB 721 ttctccatatatccagccacactcatttttaattatttagttccagatctgtactgtgac 780
QY 781 ctttctacactgtagaataaacattactcatttctgttcaaa 820
DB 781 ctttctacactgtagaataaacattactcatttctgttcaaa 820
```

RESULT 3

```
AAA06690/c
ID AAA06690 standard; cDNA; 812 BP.
XX
AC AAA06690;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:471.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
```

```
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
XX WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
XX comprises an immunogenic portion of prostate tumor protein -
PS Claim 1; Page 262; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AA082000 to AA082020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 812 BP; 198 A; 189 C; 169 G; 256 T; 0 other;

Query Match 99.0%; Score 812; DB 21; Length 812;
Best Local Similarity 100.0%; Pred. No. 1.le-246;
Matches 812; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ctggcatcagaaaaacagagggagatttgtgtgctgcagccgagggagacagaaaga 65
DB 812 CTGGCATCAGAAAAACAGAGGGGAGATTGTGTGGCTGCAGCCGAGGAGACGAGGA 753
QY 66 tctgcatggtgggaaggacctgatgacagaggtgagaaataagaaggctgctgactt 125
DB 752 TCTGCATGTGGGAAGGACCTGATATACAGAGGTGAGAAATAAGAAAGCTGCTGACTT 593
QY 126 taccatctgagggccacacatctgctgaaatggagataatataacatcactagaaacagcaa 185
DB 692 TACCATCTGAGGCCACACATCTGCTGAATGGAGATAATTAACATCATTAGAACAGCAA 633
QY 186 gatgacaataataatgtcttaagttagtagacatgtttttgcacatttccagcccttaataa 245
DB 632 GATGACAATATATGTTCTTAAGTAGTACATGTGTTTTCACATTTCCAGCCCTTTAAATA 573
QY 246 tccacacacaggaagcacaagaagcacagagatccctgggagaaatcccgccgcg 305
DB 572 TCCACACACAGGAAGACAAAAGGAAGACAGAGATCCCTGGGGAAGAAATGCCCGGCCG 513
QY 306 ccattcttggtcatgatgagcctgcctgtgcctgttcccgcttctgtgagggaagaca 365
DB 512 CCATCTTGGTTCATCGATGAGCCTCGCCCTGTGCCCTGTCCGCTGTGTGAGGAGGACA 453
QY 366 ttgaaaaatgaattgatgttctcttaagagatggcagagaaacacagatcctgtgtgga 425
DB 452 TTAGAAAATGAATTGATGTGTTCCTTTAAAGGATGGGCAGGAAACAGATCCTTGTGTGGA 393
```

QY		426	tatttatttgaaacgaggatttacagattgtaaatgaagtcaacaagtgcattaccaatga	485
Db				
		392	TATTTTATTTGAACGGGATTACAGATTGTAATGAAGTCACAAAGTGAGCATTTACCAATGA	333
QY		486	gaggaaaacagacagaaaaatcttgatggcttcacaagaacatgcacaacaaaaaatgaa	545
Db				
		332	GAGGAAAACAGACGAGAAAATCTTGTATGGCTTTCACAAGACATGCACAAACAAAAATGGAA	273
QY		546	tactgtgatgacatgaggcagccaaactcgggggaggagataaacacaggggcagaggtcacg	605
Db				
		272	TACTGTGATGACATGAGGACGCCAAGCTGGGGAGGAGATAAACCCGGGCAGGGGTCAg	213
QY		606	gattctggcccctgcctcctaactctgcgttcataaccacaaatcatctcatattctaaccc	665
Db				
		212	GATTCGGGCCCTCCTCGCTTAACACTGTCGTTTCATACCAAATCATTTCTATATTTCTAACCC	153
QY		666	ctcaaaacaaagctgttgtaatatctgatctctacggttcctctctgggcccaacattcttc	725
Db				
		152	CTCAAAACAAAGCTGTGTGAATATCTGATCTCTACGGTTCCTTCGGGGCCCAACATTCTC	93
QY		726	cataatccagccacactcaatttttaatttagttcccgatctctgactgtgaccttttc	785
Db				
		92	CATATATCCAGCCACACTCATTTTTTAATTTAGTTTCCCAGACTCTGTACTGTGACCTTTC	33
QY		786	tacactgtagaataaacattactacttttgttc	817
Db				
		32	TACACTGTAGAATAAACATTACTCATTTTTTCTC	1

RESULT	4
AAA06687	
ID	AAA06687 standard; cDNA; 3112 BP.
XX	
XX	
AC	AAA06687;
AC	
XX	
XX	13-JUN-2000 (first entry)
DT	
XX	
XX	
DE	Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:468.
XX	
XX	
KW	Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW	immunogenic; cytostatic; vaccine; ss.

XX	Homo sapiens.	
OS	WO200004149-A2.	
XX		
XX		
PN	27-JAN-2000.	
XX		
XX		
XX	14-JUL-1999;	99WO-US15838.
PF		
XX		
XX	14-JUL-1998;	98US-0115453.
PR		
PR	14-JUL-1998;	98US-0116134.
PR	23-SEP-1998;	98US-0159812.
PR	23-SEP-1998;	98US-0159822.
PR	15-JAN-1999;	99US-0232149.
PR	15-JAN-1999;	99US-0232880.
PR	09-APR-1999;	99US-0288946.
XX		
XX	(CORI-) CORIXA CORP.	
PA		
XX		
XX	Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;	
PI		
PI		
PI		
XX	WPI; 2000-171268/15.	
DR		
XX		
PT	New polypeptide useful for treating and diagnosing prostate cancer	
PT	comprises an immunogenic portion of prostate tumor protein -	
XX		
PS	Claim 1; Page 259-260; 263pp; English.	
XX		
XX		
CC	The present invention describes isolated polypeptides, comprising an	
CC	immunogenic portion of a prostate tumour protein (PTP). The polypeptides	
CC	and polynucleotides encoding them have cytostatic activity and can be	
CC	used for the treatment of prostate cancer.	

used in vaccines and in gene therapy. The polypeptides and polynucleotides encoding them, antigen presenting cells which express the polypeptides, antibodies against the polypeptides and vaccines comprising them can be used for inhibiting the development of prostate cancer in a patient. The polypeptides can be used to generate antibodies or anti-idiotypic antibodies for passive immuno therapy. A portion of the polynucleotides encoding the polypeptides can be used as a probe or to modulate the expression of the polypeptides. AAA06241 to AAA06691 and AA82000 to AA82020 represent sequences used in the exemplification of the present invention.

Query Match.	88.5%;	Score 725.8;	DB 21;	Length 3112;	
Best Local Similarity	99.7%;	Pred. No. 4.5e-219;			
Matches 727;	Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	92	tacagaggtgagaataagaagaagctgctgactttaccatctgagggcacacatctgctg	151		
DB	1307	taataaggtgagaataagaagaagctgctgactttaccatctgagggcacacatctgctg	1366		
QY	152	aaatggagataaattaaacatcactagaacagcaagatgacataataatgtctaaagtatg	211		
DB	1367	aaatggagataaattaaacatcactagaacagcaagatgacataataatgtctaaagtatg	1426		
QY	212	acatgtttttgcacatttcacagccctttaaataatccacacacagagaacacaaagg	271		
DB	1427	acatgtttttgcacatttcacagccctttaaataatccacacacagagaacacaaagg	1486		
QY	272	aagcacagagatccctgggagaaaatgccggccgccatcttgggtcatcgtatgagctcg	331		
DB	1487	aagcacagagatccctgggagaaaatgccggccgccatcttgggtcatcgtatgagctcg	1546		
QY	332	ccctgtgcctgggtccccccttgttgagggaaagacattagaaaaatgaattgatgtgttcctt	391		
DB	1547	ccctgtgcctgggtccccccttgttgagggaaagacattagaaaaatgaattgatgtgttcctt	1606		
QY	392	aaagatggggcaggaaaacagatccctgtgtggatatattttgaacgggattacagatt	451		
DB	1607	aaagatggggcaggaaaacagatccctgtgtggatatattttgaacgggattacagatt	1666		
QY	452	tgaataaagtccaacaaagttagcataccatagagagaaaacagacagaaaaatcttga	511		
DB	1667	tgaataaagtccaacaaagttagcataccatagagagaaaacagacagaaaaatcttga	1726		
QY	512	tggcttcacaaagacatgcacaaaacaaaatggaaatactgtgatgacatgaggcgccaag	571		
DB	1727	tggcttcacaaagacatgcacaaaacaaaatggaaatactgtgatgacatgaggcgccaag	1786		
QY	572	ctggggagagagataaacacaggggcagagggctcaggattcttggccctgctgcctaaactgt	631		
DB	1787	ctggggagagagataaacacaggggcagagggctcaggattcttggccctgctgcctaaactgt	1846		
QY	632	gcgttcataaccaaaatcatctcatatttctaaccctcaaaaacaaagtgttgttaatatct	691		
DB	1847	gcgttcataaccaaaatcatctcatatttctaaccctcaaaaacaaagtgttgttaatatct	1906		
QY	692	gattcttacaggttccctcttgggcccaacatttcccatatcatccagcacactcaatttta	751		
DB	1907	gattcttacaggttccctcttgggcccaacatttcccatatcatccagcacactcaatttta	1966		
QY	752	atatttagttccacagatctgtactgtgacctttctcacigttagaataaacattactcatt	811		
DB	1967	atatttagttccacagatctgtactgtgacctttctcacigttagaataaacattactcatt	2026		
QY	812	ttgttcaaaa	820		
DB	2027	ttgttcaaaa	2035		

RESULT 5
AAA06688/C

XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
XX
PS Claim 1; Page 261-262; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytotstatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 2426 BP; 717 A; 476 C; 548 G; 685 T; 0 other;

Query Match 88.5%; Score 725.6; DB 21; Length 2426;
Best Local Similarity 97.5%; Pred. No. 4.5e-219;
Matches 737; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 65 atctgcatgtggaagacacctgataacacagaggtgagaaataaagaagctgctgaact 124
|||||
DB 1802 ATTTGCTCTCAAAATGGTTCGTATGTTTCCAGGTGAGAAATTAAGAAAGGCTGCTGACT 1743
|||||
QY 125 ttaccatctgagggccacacatctgctgaaatggagataattaacatcactagaaacagca 184
|||||
DB 1742 TTACCATCTGAGGCCACACATCTGCTGAATGGAGATAATTACATCCTAGAAACAGCA 1683
|||||
QY 185 agatgacaataatagtcttaagttagtgacatgtttttgacatcttccagccctttaaatt 244
|||||
DB 1682 AGATGACAATATATGTTCTAAGTAGTGACATGTTTGTGCACATTTCCAGCCCTTTAAAT 1623
|||||
QY 245 atccacacacaggaagacacaaaagacacagagatccctggagaaaatcccggcc 304
|||||
DB 1622 ATCCACACACAGGAAGACAAAGAAAGACAGAGATCCCTGGGAGAAATGCCGGCC 1563
|||||
QY 305 gccatctgggtcatgatgagctgcctgtgcctgtgctcccgcttggaggaaggac 364
|||||
DB 1562 GCCATCTGGGTCATCATGATGAGCTCGCCCTGTGCTGGTCCCGCTGTGAGGGAAGGAC 1503
|||||
QY 365 attagaaaatgaattgatgtgttccttaagagatgggagagaaaacagatcctgtgtgtg 424
|||||
DB 1502 ATTAGAAAATGAATTGATGTGTTTCTTAAAGGATGGGACAGAAACAGATCCTGTTGTGG 1443
|||||
QY 425 atatttattgaacgggattacagattgaaatgaagtacaaagtgcattaccattg 484
|||||
DB 1442 ATATTATTATTCACGGGATTACAGATTGGAATGAAGTCAACAAGTGAGCATTTACCAATG 1383
|||||
QY 485 agaggaacacagacgagaaatcttgatggcttcacaaagacatgcacaaacaaaatgga 544
|||||
DB 1382 AGAGAAAACAGACGAGAAAATCTGTATGCTTTCACAAGACATGCAACAAACAAAATGGA 1323
|||||
QY 545 atactgtgatgacatgaggcagcaagctgtggagagataaacacggggcagagggtca 604
|||||
DB 1322 ATACTGTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1263
|||||
QY 605 ggattctggccctgctgcctaaactgtgcgttcataaccacaaatcatcttcattcttaac 664
|||||
DB 1262 GGATTCTGGCCCTGCTGCTGAATCTGCTTCAATACCAAAATCATTTTCAATTTCTAAC 1203
|||||
QY 665 cctcaaaacaaagctgttgtaatatctgatctctacggttcctctggtggcccaattct 724
|||||
DB 1202 CCTCAAAACAAAGCTGTGTAATATCTGATCTCTACGGTTCCTTCTGGGCCCAACATCT 1143
|||||
QY 725 coatatatccagccacactcatttttaataatttagttccagatctgtactgtgaccttt 784
|||||

DB 1142 CCATATATCCAGCCACACTCAATTTTAATATTTAGTTCACAGATCTGTACTGTGACCTTT 1083
QY 785 ctacactgtagaataacattactcatttcttcaaa 820
|||||
DB 1082 CTACACTGAGAAATAACTACTCAATTTTGTTCRAA 1047
RESULT 7
AAV62427
ID AAV62427 standard; cDNA; 2037 BP.
XX
AC AAV62427;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 1.
XX
KW Prostate cancer antigen cDNA splice variant 1; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 379..534
FT /tag= a
FT /product= "PCA3 protein variant 1"
FT polyA_signal 2019..2024
FT /tag= b
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
DR P-PSDB; AAW79736.
XX
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
XX
PS Claim 3; Fig 2B-2J; 11pp; English.
XX
CC The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 1 sequence comprising of exons 1, 2, 3, 4a and
CC 4b of the PCA3 gene. The PCA3 cDNA splice variant 1 sequence,
CC isolated from a human primary prostatic tumour tissue cDNA library,
CC was found in approximately 5% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunoassay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (ant)agonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 2037 BP; 622 A; 426 C; 406 G; 575 T; 8 other;

Query Match		88.5%;	Score 725.4;	DB 19;	Length 2037;
Best Local Similarity		99.9%;	Pred. No. 4.7e-219;		
Matches 726;		Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	94	cagaggtgagaataaagaaggctgctgactttacacatctgagccacacatctgctgaa	153		
DB	259	caggggtgagaataaagaaggctgctgactttacacatctgagccacacatctgctgaa	318		
QY	154	atggagataataaacatcactagaaacagcaagatgacaataataatgtcttaagtgtgac	213		
DB	319	atggagataataaacatcactagaaacagcaagatgacaataataatgtcttaagtgtgac	378		
QY	214	atgtttttgcacatttccagcccttttaaatatccacacacaggaagcacaagaagaa	273		
DB	379	atgtttttgcacatttccagcccttttaaatatccacacacaggaagcacaagaagaa	438		
QY	274	gcacagagatccctgggagaaatgccggccgcacatcttgggtcatcgatgagcctgcgc	333		
DB	439	gcacagagatccctgggagaaatgccggccgcacatcttgggtcatcgatgagcctgcgc	498		
QY	334	ctgtgcttgcctccgctgttgagggaagacattagaaaaatgaattgatgtgttccctaa	393		
DB	499	ctgtgcttgcctccgctgttgagggaagacattagaaaaatgaattgatgtgttccctaa	558		
QY	394	aggatgggcagaaacagatcctgttggatatttttgaacgggattacagatttg	453		
DB	559	aggatgggcagaaacagatcctgttggatatttttgaacgggattacagatttg	618		
QY	454	aaatgaagtcaaaagttagcattaccatgagaggaacacagacagaaaaatcttgatg	513		
DB	619	aaatgaagtcaaaagttagcattaccatgagaggaacacagacagaaaaatcttgatg	678		
QY	514	gcttcacagacatgcacaaacaaatggaaatctgtgatgacatgagcgcaagct	573		
DB	679	gcttcacagacatgcacaaacaaatggaaatctgtgatgacatgagcgcaagct	738		
QY	574	ggggaggagataaacacaggggcagaggttcagagattctggccctgctgcctaaactgtgc	633		
DB	739	ggggaggagataaacacaggggcagaggttcagagattctggccctgctgcctaaactgtgc	798		
QY	634	gttcataaccacaaatatttcatatttcaaccctcaaaacaaagctgtgtaatatctga	693		
DB	799	gttcataaccacaaatatttcatatttcaaccctcaaaacaaagctgtgtaatatctga	858		
QY	694	tctctacggttcttctggcccaacattctccatataatccagccacacactattttaat	753		
DB	859	tctctacggttcttctggcccaacattctccatataatccagccacacactattttaat	918		
QY	754	atttagttccagatctgtactgtgacctttctacactgtagaataaacattactatttt	813		
DB	919	atttagttccagatctgtactgtgacctttctacactgtagaataaacattactatttt	978		
QY	814	gttcaaaa 820			
DB	979	gttcaaaa 985			
RESULT 8					
AAV62430					
ID	AAV62430 standard; cDNA; 3582 BP.				
XX					
AC	AAV62430;				
XX					
DT	30-DEC-1998 (first entry)				
XX					
DE	Prostate cancer antigen (PCA3) wild-type cDNA.				
XX					
KW	Prostate cancer antigen cDNA; PCA3; prostatic cancer;				
KW	PC; ds.				
XX					
OS	Homo sapiens.				
XX					
FH	Key	Location/Qualifiers			

FT	CDS	401..556			
FT		/*tag= a			
FT	polyA_signal	983..987	"PCA3 protein"		
FT		/*tag= b			
FT	polyA_signal	2041..2046			
FT		/*tag= c			
FT	polyA_signal	2597..2602			
FT		/*tag= d			
FT	polyA_signal	3494..3496			
FT		/*tag= e			
XX	WO9845420-Al.				
XX	15-OCT-1998.				
XX	09-APR-1998;	98WO-CA00346.			
XX	10-APR-1997;	97US-0041836.			
XX	(DIAG-) DIAGNOCURE INC.				
XX	Bussemakers MJG;				
XX	WPI; 1998-568347/48.				
XX	P-PSDB; AAW79738.				
XX	New nucleic acid encoding prostate cancer antigen 3 - for diagnosis, prevention and treatment of prostatic cancer				
XX	Claim 3; Fig 5B-5F; 11lpp; English.				
XX	The present sequence represents the prostate cancer antigen (PCA3) wild-type cDNA sequence comprising of exons 1, 2, 3, 4a-4d of the PCA3 gene. The invention claims for PCA3 cDNA variants and the proteins they encode. The invention also claims for antibodies against PCA3 protein. The antibodies are claimed to be useful for detecting PCA3 protein in immunoassay tests, for diagnosing, assessing and prognosing of prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed to be useful for treating PC, while determining elevated levels of PCA3 (as RNA or protein) is useful for detecting a predisposition to development of PC, e.g. in prenatal tests. Detecting PCA3 protein allows differentiation between malignant and benign prostatic disease, and the level of PCA3 expression allows correlation with the grade of tumour. PCA3 protein and its fragments are also claimed to be useful in vaccines for preventing PC; in drug screens for identifying specific (ant)agonists (potentially useful therapeutically) and for studying protein-DNA interactions.				
XX	Sequence 3582 BP; 1052 A; 788 C; 679 G; 1063 T; 0 other;				
Query Match					
Best Local Similarity 88.5%; Score 725.4; DB 19; Length 3582;					
Matches 726; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY	94	cagaggtgagaaataaagaaggctgctgactttaccattctgagggccacacatctgctgaa	153		
DB	281	caggggtgagaaataaagaaggctgctgactttaccattctgagggccacacatctgctgaa	340		
QY	154	atggagataataaacatcactagaaacagcaagatgacaataataatgtcttaagtagtac	213		
DB	341	atggagataataaacatcactagaaacagcaagatgacaataataatgtcttaagtagtac	400		
QY	214	atgtttttgcacatttccagccctttaaataatccacacacaggaagcacaagaagaa	273		
DB	401	atgtttttgcacatttccagccctttaaataatccacacacaggaagcacaagaagaa	460		
QY	274	gcacagagatccctgggagaaatgccggccgcacatcttgggtcatcgatgagcctcgcc	333		
DB	461	gcacagagatccctgggagaaatgccggccgcacatcttgggtcatcgatgagcctcgcc	520		


```
QY 777 tgacctttctacactgtagaataacattactactctatttggcca 820
|||||
Db 534 tgacctttctacactgtagaataacattactactctatttggcca 577
|||||

RESULT 10
AAA06545
ID AAA06545 standard; cDNA; 718 BP.
XX
AC AAA06545;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:313.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 199-200; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AA82000 to AA82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 718 BP; 222 A; 145 C; 169 G; 172 T; 10 other;

Query Match 62.6%; Score 513.2; DB 21; Length 718;
Best Local Similarity 97.3%; Pred. No. 3.4e-152;
Matches 585; Conservative 0; Mismatches 9; Indels 7; Gaps 6;

QY 27 ggagattgtgtg-cgcagccggaggagaccaggaagatctgcatgtgtggaaggacc 85
|||||
Db 1 ggagattgtgtgttgcagccggaggagaccaggaagatctgcatgtgtggaaggacc 60
|||||
QY 86 tgatgatacagaggtgagaaataagaaaggctgctgactttaccatctgaggccacacat 145
|||||
```

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 SQ Sequence 437 BP; 140 A; 87 C; 118 G; 88 T; 4 other;

Query Match 50.5%; Score 414.2; DB 21; Length 437;
 Best Local Similarity 99.0%; Pred. No. 5.2e-121;
 Matches 413; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 agaagctgcacatacaaaaacagagggagattgtgtgctgcagcccgaggagaccag 60
 |||||
 Db 21 agaagctgcacatacaaaaacagagggagattgtgtgctgcagcccgaggagaccag 80
 |||||

QY 61 gaagatctcatgtgtggaagacacctgatgacagaggtgagaaataagaagcctgt 120
 |||||
 Db 81 gaagatctcatgtgtggaagacacctgatgacagaggtgagaaataagaagcctgt 140
 |||||

QY 121 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaaac 180
 |||||
 Db 141 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaaac 200
 |||||

QY 181 agcaagatgacataataatgtctaagtagtagacatgtttttcacattccagcccttt 240
 |||||
 Db 201 agcaagatgacataataatgtctaagtagtagacatgtttttcacattccagcccttt 260
 |||||

QY 241 aaatatccacacacacaggaagacacaaaaggagcacagagatccctgggagaaatgcc 300
 |||||
 Db 261 aaatatccacacacacaggaagacacaaaaggagcacagagatccctgggagaaatgcc 320
 |||||

QY 301 ggcgcgcattctgggtcatcgatgagcctcgccctgtgctggtcccgcttgtgagggaa 360
 |||||
 Db 321 ggcgcgcattctgggtcatcgatgagcctcgccctgtgctggtcccgcttgtgagggaa 380
 |||||

QY 361 gacattagaaatgaattgattgttccctaaaggatggcgaggaaaacagatcct 417
 |||||
 Db 381 gacattagaaatgrattgattgttccctaaaggatggcgaggaaaacagatcct 437
 |||||

RESULT 12

AAZ33445

ID AAZ33445 standard; cDNA; 359 BP.

XX

AC AAZ33445;

XX

DT 08-DEC-1999 (first entry)

XX

DE Human prostate cancer-associated EST 23.

XX

KW Expressed sequence tag; EST; prostate tumor; antitumor; treatment;

XX

KW gene therapy; tissue specificity human; ss.

XX

OS Homo sapiens.

XX

PN DE19811193-A1.

XX

PD 16-SEP-1999.

XX

PF 10-MAR-1998; 98DE-1011193.

XX

PR 10-MAR-1998; 98DE-1011193.
 XX
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX
 PI Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosenthal A;
 XX
 DR WPI; 1999-519628/44.
 DR P-PSDB; AAY48243.
 XX
 PT New nucleic acid expressed at high level in prostatic tumor tissue and
 PT encoded polypeptides, useful for treating cancer and screening for
 PT therapeutic agents -
 XX
 PS Claim 1a; 87; 166pp; German.
 XX
 CC This invention describes novel nucleic acid sequences (A) that are
 CC expressed at high level in prostatic tumor tissue and encode gene
 CC products or their fragments. The products of the invention have
 CC antitumor activity. Polypeptides (I) encoded by (A) are used: (i) for
 CC identifying agents for treatment of prostatic cancer and (ii) for
 CC therapy of prostate cancer, optionally where expressed by gene therapy
 CC methods. (A) is also used to isolate full-length genes (for gene therapy)
 CC and for recombinant production of (I), which can be used to raise
 CC specific antibodies. (A) are identified by assembly of ESTs (expressed
 CC sequence tags) before they are analyzed for expression pattern (tissue
 CC specificity). This approach eliminates many of the false results, as
 CC regards tissue specificity, associated with known methods that use
 CC single (usually short) ESTs. AAZ33423-233476 represent expressed
 CC sequence tags described in the method of the invention.
 XX
 SQ Sequence 359 BP; 121 A; 75 C; 94 G; 69 T; 0 other;

Query Match 40.2%; Score 330; DB 20; Length 359;
 Best Local Similarity 100.0%; Pred. No. 2e-94;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaagctgcacatacaaaaacagagggagattgtgtgctgcagcccgaggagaccag 60
 |||||
 Db 30 agaagctgcacatacaaaaacagagggagattgtgtgctgcagcccgaggagaccag 89
 |||||

QY 61 gaagatctcatgtgtggaagacacctgatgacagaggtgagaaataagaagcctgt 120
 |||||
 Db 90 gaagatctcatgtgtggaagacacctgatgacagaggtgagaaataagaagcctgt 149
 |||||

QY 121 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaaac 180
 |||||
 Db 150 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaaac 209
 |||||

QY 181 agcaagatgacataataatgtctaagtagtagacatgtttttcacattccagcccttt 240
 |||||
 Db 210 agcaagatgacataataatgtctaagtagtagacatgtttttcacattccagcccttt 269
 |||||

QY 241 aaatatccacacacacaggaagacacaaaaggagcacagagatccctgggagaaatgcc 300
 |||||
 Db 270 aaatatccacacacacaggaagacacaaaaggagcacagagatccctgggagaaatgcc 329
 |||||

QY 301 ggcgcgcattctgggtcatcgatgagcctc 330
 |||||
 Db 330 ggcgcgcattctgggtcatcgatgagcctc 359
 |||||

RESULT 13

AAA06520/c

ID AAA06520 standard; cDNA; 301 BP.

XX

AC AAA06520;

XX

DT 13-JUN-2000 (first entry)

XX

DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:287.

XX

KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;

immunogenic; cytostatic; vaccine; ss.
Homo sapiens.
WO200004149-A2.
27-JAN-2000.
14-JUL-1999; 99WO-US15838.
14-JUL-1998; 98US-0115453.
14-JUL-1998; 98US-0116134.
23-SEP-1998; 98US-01159812.
23-SEP-1998; 98US-01159822.
15-JAN-1999; 98US-0232149.
15-JAN-1999; 99US-0232880.
09-APR-1999; 99US-0288946.
(CORI-) CORIXA CORP.
Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
WPI; 2000-171268/15.
New polypeptide useful for treating and diagnosing prostate cancer
comprises an immunogenic portion of prostate tumor protein -
Claim 1; Page 192; 263pp; English.
The present invention describes isolated polypeptides, comprising an
immunogenic portion of a prostate tumour protein (pTP). The polypeptides
and polynucleotides encoding them have cytostatic activity and can be
used in vaccines and in gene therapy. The polypeptides and
polynucleotides encoding them, antigen presenting cells which express
the polypeptides, antibodies against the polypeptides and vaccines
comprising them can be used for inhibiting the development of prostate
cancer in a patient. The polypeptides can be used to generate antibodies
or anti-idiotypic antibodies for passive immuno therapy. A portion of
the polynucleotides encoding the polypeptides can be used as a probe or
to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
AAV82000 to AAV82020 represent sequences used in the exemplification of
the present invention.
Sequence 301 BP; 76 A; 58 C; 70 G; 97 T; 0 other;

Query Match 35.2%; Score 288.4; DB 21; Length 301;
Best Local Similarity 99.3%; Pred. No. 2.6e-81;
Matches 300; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
Qy 472 agcattaccatgagagaaacagacagagaaatcttgatggtctcacaagacatgcaa 531
Db 301 AGCATTACCATGAGAGAAACAGACAGAGAAATCTTGATGGCTTCACAGACATGCAA 242
Qy 532 caaacaaaatggaatactgtgatgacatgagcgagcgaagctggggagagataaccacg 591
Db 241 CAACAAAATGGAATGTAATGTAACATGAGCGAGCCAAAGCTGGGGAGGAGATAACACG 182
Qy 592 gggcagaggggcaggattctggcctgctgctaaactgtcggttcataaccataatcatt 651
Db 181 GGGCAGAGGGGTCAGGATTTCTGGCCCTGCTGCTAAACGTGCTTCAATACCAATCAT 122
Qy 652 tcatattctaacctcacaacaaagctgtgttaatactgactctcactcctctctg 711
Db 121 TCATATTTCTAACCTCAAAACAAAGCTGTGTATATATCTGATCTCTAC-GTTCCTCTG 63
Qy 712 gggcccaacttctccatatccagccacactcatttttaattagttccagatctg 771
Db 62 GGCCCAACATCTCCATATATCCAGCCACACTCATTTTTATATATTAGTTCCTCCAGATCTG 3
Qy 772 ta 773
Db 2 TA 1

RESULT 14
AAV33791
ID AAV33791 standard; cDNA; 123 BP.
XX AC AAV33791;
XX 30-DEC-1998 (first entry)
XX Prostate cancer antigen (PCA3) gene transcription start site.
XX Prostate cancer antigen gene transcription start site; PCA3;
XX prostatic cancer; PC; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FT misc_feature 122..123
FT /*tag= a
FT /note= "Splice acceptor site"
XX WO9845420-A1.
XX 15-OCT-1998.
XX 09-APR-1998; 98WO-CA00346.
XX 10-APR-1997; 97US-0041836.
XX (DIAG-) DIAGNOCURE INC.
XX Bussemakers MJG;
XX WPI; 1998-568347/48.
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
XX prevention and treatment of prostatic cancer
XX Example 3; Fig 4; 11pp; English.
XX The present sequence represents a transcription start site sequence
XX of the prostate cancer antigen (PCA3) gene. The invention claims for
XX PCA3 cDNA variants and the proteins they encode. The invention also
XX claims for antibodies against PCA3 protein. The antibodies are
XX claimed to be useful for detecting PCA3 protein in immunoassay tests,
XX for diagnosing, assessing and prognosing of prostatic cancer (PC).
XX Antibodies, optionally coupled to a cytotoxin or radioisotope, and
XX nucleic acids antisense to PCA3 cDNA are claimed to be useful for
XX treating PC, while determining elevated levels of PCA3 (as RNA or
XX protein) is useful for detecting a predisposition to development of
XX PC, e.g. in prenatal tests. Detecting PCA3 protein allows
XX differentiation between malignant and benign prostatic disease,
XX and the level of PCA3 expression allows correlation with the grade of
XX tumour. PCA3 protein and its fragments are also claimed to be useful
XX in vaccines for preventing PC; in drug screens for identifying
XX specific (ant)agonists (potentially useful therapeutically) and for
XX studying protein-DNA interactions.
XX Sequence 123 BP; 41 A; 19 C; 45 G; 18 T; 0 other;

Query Match 12.2%; Score 100; DB 19; Length 123;
Best Local Similarity 100.0%; Pred. No. 6.1e-22;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 agaagctggcatcagaaaaacagagggagagatttggctgcagccgagggagaccag 60
Db 24 agaagctggcatcagaaaaacagagggagagatttggctgcagccgagggagaccag 83
Qy 61 gaagatctgcatggtgggaagaccctgatgatacagagt 100
Db 84 gaagatctgcatggtgggaagaccctgatgatacagagt 123

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 08:05:28 ; Search time 10902.7 Seconds
(without alignments)
1163.342 Million cell updates/sec

Title: US-09-402-713A-4
Perfect score: 820
Sequence: 1 agaagctgcatcagaaaaa.....cattactcattgttcaaa 820

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_bal.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pl1.*
13: gb_pl2.*
14: gb_pl3.*
15: gb_pl4.*
16: gb_bal.*
17: gb_ba2.*
18: em_fun.*
19: em_htgo_hum.*
20: em_htgo_inv.*
21: em_htgo_rod.*
22: em_htg_hum1.*
23: em_htg_hum2.*
24: em_htg_hum3.*
25: em_htg_hum4.*
26: em_htg_hum5.*
27: em_htg_hum6.*
28: em_htg_hum7.*
29: em_htg_hum8.*
30: em_htg_inv1.*
31: em_htg_inv2.*
32: em_htg_other.*
33: em_htg_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_vl.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_vl1.*
59: gb_vl2.*
60: gb_htg1.*
61: gb_htg2.*
62: gb_htg3.*
63: gb_htg4.*
64: gb_htg5.*
65: gb_htg6.*
66: gb_htg7.*
67: gb_htg8.*
68: gb_htg9.*
69: gb_htg10.*
70: gb_htg11.*
71: gb_htg12.*
72: gb_htg13.*
73: gb_htg14.*
74: gb_htg15.*
75: gb_htg16.*
76: gb_htg17.*
77: gb_htg18.*
78: gb_htg19.*
79: gb_htg20.*
80: gb_htg21.*
81: gb_htg22.*
82: gb_htg23.*
83: gb_htg24.*
84: gb_htg25.*
85: gb_pr1.*
86: gb_pr2.*
87: gb_pr3.*
88: gb_pr4.*
89: gb_pr5.*
90: gb_pr6.*
91: gb_pr7.*
92: gb_pr8.*
93: gb_pr9.*
94: gb_ro1.*
95: gb_ro2.*
96: gb_in4.*
97: gb_pr10.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	725.4	88.5	3923	88	AF103907
2	542.4	66.1	5435	88	AF103908
3	542.4	66.1	164371	80	AL390239
4	542.4	66.1	173831	80	AL359314
5	542.4	66.1	267581	80	AL358573
6	330	40.2	359	9	AX018075
7	100.2	12.2	143675	79	AL161625
8	100.2	12.2	267581	80	AL358573

```

c 9 98.6 12.0 172298 61 AC009556
c 10 98 12.0 580 89 AF279290
c 11 57.4 7.0 7218 10 I66494
c 12 45 5.5 191319 63 AC015653
c 13 42.8 5.2 9131 85 AB038781
c 14 42.6 5.2 8912 85 AB038782
c 15 42.6 5.2 10708 89 AF113616
c 16 42 5.1 175144 73 AC068502
c 17 42 5.1 199103 76 AC080076
c 18 41.8 5.1 2458 94 MMSTRA1GN
c 19 40.8 5.0 115583 93 HSJ351K20
c 20 40.8 5.0 164478 83 AP002450
c 21 39.8 4.9 141492 73 AC068941
c 22 39.8 4.9 175353 73 AC069061
c 23 39.8 4.9 176958 74 AC069123
c 24 39.6 4.8 76175 15 TIN6
c 25 39.4 4.8 173547 79 AL162716
c 26 39.2 4.8 50883 90 AL359708
c 27 39 4.8 993 94 MMSNRPN3
c 28 39 4.8 123004 92 HS196823
c 29 39 4.8 175355 64 AC016480
c 30 39 4.8 183259 76 AC079914
c 31 38.8 4.7 40998 86 AC005503
c 32 38.8 4.7 41730 93 HSAC000372
c 33 38.8 4.7 129290 66 AC020761
c 34 38.8 4.7 145050 70 AC026679
c 35 38.8 4.7 178256 70 AC026878
c 36 38.8 4.7 190778 92 HS142L7
c 37 38.6 4.7 193930 86 AC006296
c 38 38.6 4.7 196607 70 AC028602
c 39 38.6 4.7 210348 69 AC025845
c 40 38.4 4.7 8569 93 HSPLPINT
c 41 38.4 4.7 17427 93 HSA6976
c 42 38.4 4.7 42013 97 HSV698D2
c 43 38.4 4.7 142084 84 HS540A13
c 44 38.4 4.7 147055 92 HS372K1
c 45 38.4 4.7 171430 70 AC027160

```

ALIGNMENTS

```

RESULT 1
AF103907
LOCUS AF103907 3923 bp mRNA PRI 14-AUG-2000
DEFINITION Homo sapiens non-coding RNA DD3 sequence.
ACCESSION AF103907
VERSION AF103907.1 GI:6165973
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3923)
AUTHORS Bussemakers,M.J., van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F., Schalken,J.A., Debruyne,F.M., Ru.N. and Isaacs,W.B.
TITLE DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072260
PUBMED 10606244
REFERENCE 2 (bases 1 to 3923)
AUTHORS Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru.N. and
Isaacs,W.B.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
Location/Qualifiers
1..3923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
source

```

```

/map="9q21-q22"
/tissue_type="prostate"
/note="non-coding RNA DD3"
repeat_region 3756..3911
/note="LINE"
/rpt_family="L1"
/rpt_type=dispersed
polya_site 3923
BASE COUNT 1157 a 840 c 740 g 1186 t
ORIGIN
Query Match 88.5%; Score 725.4; DB 88; Length 3923;
Best Local Similarity 99.9%; Pred. No. 5.8e-195;
Matches 726; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 94 cagaggtgagaaataagaaggctgtgtgactttacacatctgagccacacatctgtgtgaa 153
DB 281 CAGGGGTGAGAAATAGAAAGGCTGCTGACTTTTACCATCTGAGGCCACACATCTGCTGAA 340
QY 154 atggagataatataacatcactagaaacagcaagatgacaataataatgtctaaagtgtgac 213
DB 341 ATGGAGATATTAACATCACTAGAAACAGCAAGATGACAAATTAATGTCTTAAGTAGTGAC 400
QY 214 atgtttttgcacatttccagcccttttaaatatccacacacagaggaagcacaaaaggaa 273
DB 401 ATGTTTTTGCACATTTCCAGCCCTTTTAAATATCCACACACAGAGGAAGCACAAGAGGAA 460
QY 274 gaacagagatccctgggagaaatgcgcgcgcctctgtgggtcactgtagcctgcgc 333
DB 461 GCACAGAGATCCCTGGGAGAAATGCCCGCGCCATCTTTGGGTCAATCGATGAGCCTCGCC 520
QY 334 ctgtgcctgtccctctgtgagggaagacattagaaaaatgaatgattgttctcttaa 393
DB 521 CTGTGCTGCTCCCGCTTGTGAGGGAAGGACATAGAAAATGAATGTGTGTCTCTTAA 580
QY 394 aggatgggagaaacacagatcctgtgtgggatatttatttgacgggattacagatttg 453
DB 581 AGGATGGCAGGAGAAACAGATCCTGTTGTGGATATTTATTTGAACGGGATTACAGATTTG 640
QY 454 aaatgaagtcacaaagtgcattaccatgagaggaagaaacagagaaaaatcttgatg 513
DB 641 AAATGAAGTCACAAAGTGAGCATTACCAATGAGAGGAGAAACAGAGAGAAAATCTTGATG 700
QY 514 gcttcacagacatgcaacaaacaaatactgtgatactgtgacatgagggcagccaagct 573
DB 701 GCTTCACAGACATGCAACAAACAAATGAATGGAATGCTGTGATGATGAGGCACCAAGCT 760
QY 574 ggggaggagataaccacggggcagagggtcaggattcttgccctgctgctaaactgtgc 633
DB 761 GGGGAGGAGATAACACACGGGGCAGAGGGTCAGGATTTCTGGCCCTGCTCTAACTGTC 820
QY 634 gttcataacaaataatcttcatatttctaacctcacaacaaagctgtgtgataatctga 693
DB 821 GTTCATAACCAATCATATTTTCATATTTTAACTTAACTTAACTTAACTTAACTTAACTTAA 880
QY 694 tctctacgattcttctggcccaacatctcattatccacacacacacacacacacacacac 753
DB 881 TCCTACGGTTCCTCTGGGCCCAACATCTCTCATATATCCACCCACACACTCATTTTAAAT 940
QY 754 attagttccagatctgtactgtgacctttctacactgtagataacattactatctt 813
DB 941 ATTAGTTCACAGATCTGTACTGTGACCTTTCTACACTGTAGATAACATTACTCATTTT 1000
QY 814 gttcaaaa 820
DB 1001 GTTCAAA 1007
RESULT 2
AF103908
LOCUS AF103908 5435 bp DNA PRI 14-AUG-2000
DEFINITION Homo sapiens non-coding RNA DD3 gene, exons 2, 3, and 4.

```

AF103908
 VERSION AF103908.1 GI:6165974
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 5435)
 AUTHORS Bussemakers, M.J., van Bokhoven, A., Verhaegh, G.W., Smit, F.P.,
 Karthaus, H.F., Schalken, J.A., Debruyne, F.M., Ru, N. and Isaacs, W.B.
 TITLE DD3: a new prostate-specific gene, highly overexpressed in prostate
 cancer
 JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
 MEDLINE 20072260
 PUBMED 10606244
 REFERENCE 2 (bases 1 to 5435)
 AUTHORS Bussemakers, M.J.G., Van Bokhoven, A., Verhaegh, G.W., Smit, F.P.,
 Karthaus, H.F.M., Schalken, J.A., Debruyne, F.M.J., Ru, N. and
 Isaacs, W.B.
 TITLE Direct Submission
 JOURNAL Submitted (28-Oct-1998) Urology Research Laboratory, University
 Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
 FEATURES
 source
 1..5435
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /map="9q21-q22"
 1..78
 /rpt_family="Alu"
 /rpt_type="dispersed"
 533..697
 /note="alternative exon present in 5% of cDNA clones"
 /number=2
 1035..1294
 /rpt_family="Alu"
 /rpt_type="dispersed"
 join(1571..1753,1981..5435)
 /note="transcript III"
 /product="non-coding RNA DD3"
 join(1571..1753,1981..2517)
 /note="transcript I"
 /product="non-coding RNA DD3"
 join(1571..1753,1981..3579)
 /note="transcript (major) II"
 /product="non-coding RNA DD3"
 1571..1753
 /number=3
 1981..5435
 /number=4
 2495..2499
 /note="transcript I"
 2517
 /note="transcript I"
 3553..3558
 /note="transcript (major) II"
 3579
 /note="transcript (major) II"
 5268..5423
 /note="LINE"
 /rpt_family="L1"
 /rpt_type="dispersed"
 5435
 /note="transcript III"
 1546 a 1188 c 1036 g 1654 t 11 others
 BASE COUNT
 ORIGIN
 Query Match 66.1%; Score 542.4; DB 88; Length 5435;
 Best Local Similarity 99.8%; Pred. No. 5.4e-143;
 Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 277 cagagatccctgggagaaatgccgcgcgcctcttggtgcatcgatgagcctgcctg

Db 1976 CACAGATCCCTGGAGAAATGCCGCGCCCATCTTGGGTTCATCGATGAGCCTCGCCCTG 2035
 Qy 337 tgcctggctccgctgtgaggaagacattagaataatgaattgattgtcttcttaaaag 396
 Db 2036 TGCCCTGGTCCCGCTGTGTAGGGAAGGACATTAGAAAATGAATTTGATGTCTTCTTAAGG 2095
 Qy 397 atgggcagagaaacagatcctgtgtggatattttgaacgggattacagatttgaaa 456
 Db 2096 ATGGGCAGGAAAACAGATCCTGTGTGGATATTATTTCACGGGATACAGATTTGAAA 2155
 Qy 457 tgaagtacaaagttagcattaccatgagagagaaacagagagaaaaatcttgatggct 516
 Db 2156 TGAAGTCAAAAAGTGCAGCATTTACCAATGAGAGAAAACAGACAGAAAATCTTGATGGCT 2215
 Qy 517 tcacaagacatcaacaaacaaatggaatactgtgatacatgacatgagcagcaaacctgg 576
 Db 2216 TCACAGAGATCAACAAACAAATGGAATCTGTGATGACATGAGGACCAAGCTGGG 2275
 Qy 577 gaggagataaccacggggcagagggctcagattctggccctgctgcctaaactgtgcgtt 636
 Db 2276 GAGGAGATAACCCAGCGGCGAGAGGTTCAGGATTCGTGGCCCTGCTGCTAACTGTGCGTT 2335
 Qy 637 cataccaaaatcatttcatttctaaacctcaaaacaaagctgtgtgtaatatctgatct 696
 Db 2336 CATAACCAAATCATTTTCATATTCTAAACCTCAAAACAAAGCTGTTGTAATATCTGATCT 2395
 Qy 697 ctacggttcctctctggcccaacattctccatatatccacacacactcattttaatatt 756
 Db 2396 CTACGGTTCCTCTGGGCCCCAACATTCATATATCCAGCCACACTCATTTTAATATT 2455
 Qy 757 tagttccacagatctgactgtgacctttctacactgtagaataaacattactatttgtt 816
 Db 2456 TAGTTCACAGATCTGTACTGTGACCTTCTACACTGTAGAATAACATTACTATTGTGT 2515
 Qy 817 caaa 820
 Db 2516 CAAA 2519
 RESULT 3
 AL390239
 LOCUS Homo sapiens chromosome 9 clone RP11-58J3, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 4 unordered pieces.
 ACCESSION AL390239
 VERSION AL390239.11 GI:13274794
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 164371)
 AUTHORS Burton, J.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 COMMENT On Mar 12, 2001 this sequence version replaced gi:13273805.
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA58J3
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 16337 bases at least Q40
 Consensus quality: 163481 bases at least Q30
 Consensus quality: 163577 bases at least Q20

Db	33616	GAGGAGATACACCGGGGACAGGGGTGAGGATTCGGCCCTGCGCTTAAACTGTGCGTT	33673
Qy	637	cataccaataatcatttcattatatttcaaccctccaaaaaagaagctgttgtaatatctgatct	696
Db	33676	CATAACCAATCAATTCATATTTCTAAACCTCAAAACAAAGCTGTTGAATATCTGATCT	33735
Qy	697	ctacggtctctctggtggcccaacattctccatatatccagccacacactcaatttttaatt	756
Db	33736	CTACGGTTCCTTCGTGGGCCAAACATTCCTCATATATCCAGCCACACTCATATTTTAATATT	33795
Qy	757	tagttccagatctgtactgtgaactttctacactgttagaataacattactcatttggtt	816
Db	33796	TAGTTCACAGATCTGTACTGTGACCTTTCTACACTGTAGAAATAACATTACTCATTTTGT	33855
Qy	817	caaa 820	
Db	33856	CAAA 33859	
RESULT	4		
AL359314			
LOCUS	AL359314	173831 bp	DNA HTG 08-APR-2001
DEFINITION	Homo sapiens chromosome 9 clone RP11-108L4, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.		
ACCESSION	AL359314		
VERSION	AL359314.12	GI:13396560	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1. (bases 1 to 173831)		
TITLE	Leongamornlert.D.		
JOURNAL	Direct Submission		
COMMENT	Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequest@sanger.ac.uk On Mar 20, 2001 this sequence version replaced gi:13277120. ----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: bA108L4 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 173545 bases at least Q40 Consensus quality: 173599 bases at least Q30 Consensus quality: 173621 bases at least Q20 Insert size: 173631; sum-of-contigs Insert size: 172123; 10.0% error; agarose- Quality coverage: 10.35x in Q20 bases; sum-of-contigs Quality coverage: 10.50x in Q20 bases; agarose- ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 3 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 50595: contig of 50595 bp in length * 50596 50695: gap of 100 bp * 50696 157636: contig of 106941 bp in length * 157637 157736: gap of 100 bp * 157737 173831: contig of 16095 bp in length. Location/Qualifiers 1. .173831 /organism="Homo sapiens"		
FEATURES			
source			

```
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-108L4"
/clone_lib="RPC1-11.1"
1..50595
/note="assembly_fragment:04137
fragment_chain:1"
50696..157636
/note="assembly_fragment:04236
fragment_chain:1"
157737..173831
/note="assembly_fragment:01656"
52107 a 36469 c 35018 g 50037 t 200 others
ORIGIN

Query Match 66.1%; Score 542.4; DB 80; Length 173831;
Best Local Similarity 99.8%; Pred. No. 9.2e-143;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 277 cagagatccctggagaaatgccgcgccctcttgggtcatcgatgagcctcgccctg 336
|||
Db 156237 CACAGATCCCTGGGAGAAATGCCGCCGCCATCTTGGGTCTATGATGAGCCTCGCCCTG 156296

QY 337 tgctgtgtccctgtgtgagggaagacattagaaatgaattgatgtgttccctaaagg 396
|||||
Db 156297 TGCTGTGTCCCGCTTGTGAGGGAAGGACATTAGAAATGAATTGATGTCTCTTTAAAGG 156356

QY 397 atgggcagaaacagatcctgtgtggatattatttgaacgggattcacagatttgaaa 456
|||||
Db 156357 ATGGGCAGGAAACAGATCCTGTTGTGGATATTATTTGAACGGGATTACAGATTTGAAA 156416

QY 457 tgaagtcaaaagtgcattaccatgagagaaacacagacgagaaatcttgatggct 516
|||||
Db 156417 TGAAGTCACAAAGTGACATACCAATGAGAGAAACAGACGAGAAATCTTGATGGCT 156476

QY 517 tcaagacatgcacaaacaaatggaatactgtgatgacatgagcgacgaagctggg 576
|||||
Db 156477 TCACAGACATGCAACAAACAAATGGAATGATGTGATGACATGAGCGACCAAGCTGGG 156536

QY 577 gaggagataaccaggggagagggtcaggattctgtgacctgctgacctaaactgtcggt 636
|||||
Db 156537 GAGGAGATAACCAACGGGGCAGAGGGTCAGAGATTCTGGCCCTGCTGCTAAACTGTGCGTT 156596

QY 637 cataacaaatcatttcattatttcaacctcaaaacaaagctgttgatattctgatct 696
|||||
Db 156597 CATACCAATCATTTTCATATTTCTAACCTCAAAACAAGCTGTGTAATATCTGATCT 156656

QY 697 ctacggttcccttggtggcccaacattctccatatatccagccacactcatattttaaatt 756
|||||
Db 156657 CTACGTTCTCTTGTGGGCCCAACATTTCTCATATATCCAGCCACACTCATTTTAATATT 156716

QY 757 taagtccagatgtactgtgacctttacactgtagaataaacattactcatttgg 816
|||||
Db 156717 TAGTCCAGATCTGACTGTGACCTTTTACACTGTAGAATAACATACTACTATTGT 156776

QY 817 caca 820
|||||
Db 156777 CAAA 156780

RESULT 5
AL358573/c
LOCUS
DEFINITION Homo sapiens chromosome 9 clone RP11-133022, *** SEQUENCING IN
PROGRESS ***, 37 unordered pieces.
ACCESSION AL358573
VERSION AL358573.17 GI:13660951
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
AUTHORS Mclay, K.
TITLE Direct Submission
JOURNAL Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13398774.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bal33022
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 256699 bases at least Q40
Consensus quality: 259744 bases at least Q30
Consensus quality: 261407 bases at least Q20
Insert size: 263981; sum-of-contigs
Insert size: 135491; 19.3% error; agarose-fp
Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality
coverage: 12.23x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 25718: contig of 25718 bp in length
* 25719 25818: gap of 100 bp
* 25819 29567: contig of 3749 bp in length
* 29568 29667: gap of 100 bp
* 29668 40520: contig of 10853 bp in length
* 40521 40620: gap of 100 bp
* 40621 50002: contig of 9382 bp in length
* 50003 50102: gap of 100 bp
* 50103 54942: contig of 4840 bp in length
* 54943 55042: gap of 100 bp
* 55043 59067: contig of 4025 bp in length
* 59068 59167: gap of 100 bp
* 59168 72327: contig of 13160 bp in length
* 72328 72427: gap of 100 bp
* 72428 79396: contig of 6969 bp in length
* 79397 79496: gap of 100 bp
* 79497 82738: contig of 3242 bp in length
* 82739 82838: gap of 100 bp
* 82839 87543: contig of 4705 bp in length
* 87544 87643: gap of 100 bp
* 87644 89487: contig of 2204 bp in length
* 89488 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7323 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length
* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2383 bp in length
* 121870 121969: gap of 100 bp
* 121970 125546: contig of 3577 bp in length
*
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

* 125547 125646: gap of 100 bp
* 125647 128990: contig of 334 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length
* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2198 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.

FEATURES

Source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-133022"
/clone_lib="RP11-11.1"
1..25718
/note="assembly_fragment:00645
fragment_chain:1
clone_end:T7
vector_side:left"
25819..29567
/note="assembly_fragment:01660
fragment_chain:1"
29668..40520
/note="assembly_fragment:03080
fragment_chain:1"
40621..50002
/note="assembly_fragment:00771
fragment_chain:1"
50103..54942
/note="assembly_fragment:02629
fragment_chain:1"
55043..59067
/note="assembly_fragment:04565
fragment_chain:1"
59168..72327
/note="assembly_fragment:00223
fragment_chain:1"
72428..79396
/note="assembly_fragment:03318
fragment_chain:1"
79497..82738
/note="assembly_fragment:00224
fragment_chain:2"
82839..87543
/note="assembly_fragment:03050
fragment_chain:2"

misc_feature
87644..89847
/note="assembly_fragment:01948
fragment_chain:3"
89948..92376
/note="assembly_fragment:02191
fragment_chain:3"
92477..99799
/note="assembly_fragment:00010"
99900..102117
/note="assembly_fragment:00672"
102218..109330
/note="assembly_fragment:01326"
109431..112187
/note="assembly_fragment:01498"
112288..116407
/note="assembly_fragment:01607"
116508..119386
/note="assembly_fragment:01899"
119487..121869
/note="assembly_fragment:02068"
121970..125546
/note="assembly_fragment:02469"
125647..128990
/note="assembly_fragment:02502"
129091..134055
/note="assembly_fragment:02669"
134156..138314
/note="assembly_fragment:02757"
138415..140612
/note="assembly_fragment:02776"
140713..143296
/note="assembly_fragment:03127"
143397..148113
/note="assembly_fragment:03634"
148214..151009
/note="assembly_fragment:03898"
151110..156616
/note="assembly_fragment:04220"
156717..158763
/note="assembly_fragment:04276"
158864..162129
/note="assembly_fragment:04418"
162230..164443
/note="assembly_fragment:04428"

Query Match 66.1%; Score 542.4; DB 80; Length 267581;
Best Local Similarity 99.8%; Pred.No.9.8e-143;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 277 cagagatccctgggagaaatgccgcgcgcctcttgggtcatcgatgagcctcgccctg 336
|||
Db 165900 CACAGATCCCTGGGAGAAATGCCGCCGCCCATCTGGGTCTATGATGAGCTCGCCCTG 165841
QY 337 tgcctggctccctgtgtgaggaagacattagaaaaatgaattgtgttccctaaagg 396
|||
Db 165840 TGCCTGCTCCCTGTGTGAGGGAAGACATTAGAAAATGAATTGATGTCTCTTAAAGG 165781
QY 397 atggcagagaaacagatccttattgatatatttttaacggaattacatttga 456
|||
Db 165780 ATGGCGAGGAAACAGATCCTGTGTGGATATTTATTTGAACGGGATACAGATTGAAA 165721
QY 457 tgaagtcaacaagtgcgcatattaccatgagagagaaacagacagaaaaatccttgagct 516
|||
Db 165720 TGAAGTCACAAAGTGAGCATACCAATGAGAGGAAACACAGAGAAAATCTTGATGGCT 165661
QY 517 tcacaagacatgcaacaacaaatggaatactgtgatgacatgagcagcaagctggg 576
|||
Db 165660 TCACAAGACATGCAACAAAATGGAATACTGTGATGACATGAGGACCAAGCTGGG 165601
QY 577 gaggagataacacaggggcagaggttcaggattctggccctctgcctaaactgtcgctt 636
|||
Db 165600 GAGGAGATATACCGGGGCAGAGGGTCTGCTGCTTAACTGTGCGTT 165541

```
QY 637 catacaaaatcatttcatttctaaacctaaacaaagctgttgtaattctgatct 696
|||||
Db 165540 CATAACCAATCATTTTCATATTTCTAAACCTCAAAACAAGCTGTGTATATCTGATCT 165481
|||||
QY 697 ctacggctctcttgggcccaacattctccatatatccacacactcatttttaatt 756
|||||
Db 165480 CTACGGTCTCTTCTGGGCCCAACATTTCTCATATATCCAGCCACACTCATTTTAAAT 165421
|||||
QY 757 tagttccagatctgactgtgacctttctacactgtagaataacattactcattttgt 816
|||||
Db 165420 TAGTTCACAGATCTGACTGTGACCTTTCTACACTAGATAAACAATTACTCATTTTGT 165361
|||||
QY 817 caaa 820
|||||
Db 165360 CAAA 165357

RESULT 6
AX018075
LOCUS AX018075 359 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 23 from Patent W09946374.
ACCESSION AX018075
VERSION AX018075.1 GI:10042526
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 359)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
TITLE Human nucleic acid sequences from prostate tumour tissue
JOURNAL Patent: WO 9946374-A 23 16-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
source
1..359
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 121 a 75 c 94 g 69 t
ORIGIN

Query Match 40.2%; Score 330; DB 9; Length 359;
Best Local Similarity 100.0%; Pred. No. 7.le-83;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaagctggcatcagaaaaacagagggagatttggctgcagccgagggagaccag 60
|||||
Db 30 AGAAGCTGGCATCAGAAAAACAGAGGGGAGATTGTGTGGCTGCAGCCGAGGGAGACCAG 89
|||||
QY 61 gaagatctgcattggtgggaagacctgatgatacagagtgagaaataagaaggctgt 120
|||||
Db 90 GAAGATCTGCATTGTTGGGGAAGGACCTGTGATACACAGGTGAGAAATAAGAAAGGCTGT 149
|||||
QY 121 gactttaccatctgagccacacatctctgaaatggagataattaacatcactagaac 180
|||||
Db 150 GACTTTACCATCTGAGGCCACACATCTGCTGAATGGAGATTAATTAACATCACTAGAAC 209
|||||
QY 181 agcaagatgacataaatgtctaagtagtgacatgttttgcacatttccagcccttt 240
|||||
Db 210 AGCAAGATGACATAATAGTCTAAGTAGTGACATGTTTTGCACATTTCCAGCCCTTT 269
|||||
QY 241 aaatataccacacacagaaagcacaaaaaggaagcacagagatccctgggagaaaagccc 300
|||||
Db 270 AAATATCCACACACAGGAAGCACAAAAGGAAGCACAGAGATCCCTGGGAGAAATGCC 329
|||||
QY 301 ggcgcgcatcttggtcatcgatgagcctc 330
|||||
Db 330 GGCCCGCATCTTTGGGTGTCATCGATGAGCCCTC 359
|||||
```

```
RESULT 7
AL161625
LOCUS AL161625 143675 bp DNA HTG 20-JAN-2001
DEFINITION Homo sapiens chromosome 9 clone RP11-146P9, *** SEQUENCING IN
PROGRESS ***, 10 unordered pieces.
ACCESSION AL161625
VERSION AL161625.6 GI:9863607
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143675)
AUTHORS Plumb,B.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:8894260.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bA146P9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 138647 bases at least Q40
Consensus quality: 140619 bases at least Q30
Consensus quality: 141706 bases at least Q20
Insert size: 142775; sum-of-contigs
Insert size: 147523; 5.9% error; agarose-fp
Quality coverage: 4.29x in Q20 bases; sum-of-contigs Quality
coverage: 4.22x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 12163: contig of 12163 bp in length
* 12164 12263: gap of 100 bp
* 12264 25269: contig of 13006 bp in length
* 25270 25369: gap of 100 bp
* 25370 57709: contig of 32340 bp in length
* 57710 57809: gap of 100 bp
* 57810 66792: contig of 8983 bp in length
* 66793 66892: gap of 100 bp
* 66893 102772: contig of 35880 bp in length
* 102773 102872: gap of 100 bp
* 102873 106863: contig of 3991 bp in length
* 106864 106963: gap of 100 bp
* 106964 129804: contig of 22841 bp in length
* 129805 129904: gap of 100 bp
* 129905 132223: contig of 2319 bp in length
* 132224 132323: gap of 100 bp
* 132324 135764: contig of 3441 bp in length
* 135765 135864: gap of 100 bp
* 135865 143675: contig of 7811 bp in length.
FEATURES
Location/Qualifiers
1..143675
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-146P9"
/clone_lib="RPCI-11.1"
misc_feature 1..12163
```

/note="assembly_fragment:00440
clone_end:77
vector_side:left"
misc_feature 12264..25269
/note="assembly_fragment:00434
fragment_chain:1"
misc_feature 25370..57709
/note="assembly_fragment:00526
fragment_chain:1"
misc_feature 57810..66792
/note="assembly_fragment:00661
fragment_chain:1"
misc_feature 66893..102772
/note="assembly_fragment:00323"
102873..106863
/note="assembly_fragment:00860"
106964..129804
/note="assembly_fragment:00978"
129905..132223
/note="assembly_fragment:01068"
132324..135764
/note="assembly_fragment:01424"
135865..143675
/note="assembly_fragment:01413
clone_end:SP6
vector_side:right"
BASE COUNT 42329 a 30042 c 28964 g 41431 t 909 others
ORIGIN

Query Match 12.2%; Score 100.2; DB 79; Length 143675;
Best Local Similarity 97.1%; Pred. No. 3e-17;
Matches 102; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 agaagctgcacagaaaacagagggagattgtgtgctgcacagccagagagacag 60
Db 136065 AGAAGCTGCATCAGAAAACAGAGGGGAGATTGTGTGCTGCACGCCGAGGAGACCAG 136124
QY 61 gaagatctgcattgtgggaagccatgatcatcagaggtgagaa 105
Db 136125 GAAGATCTGCATGTTGGGAGGACCTGATCATACAGAGTCTGTA 136169

RESULT 8
AL358573
LOCUS
DEFINITION Homo sapiens chromosome 9 clone RP11-133022, HTG 15-APR-2001
PROGRESS ***, 37 unordered pieces.
ACCESSION AL358573
VERSION AL358573.17 GI:13660951
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 267581)
Mclay,K.
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13398774.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bal33022
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 256699 bases at least Q40
Consensus quality: 259744 bases at least Q30
Consensus quality: 261407 bases at least Q20
Insert size: 263981; sum-of-contigs
Insert size: 135491; 19.3% error; agarose-fp
Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality
coverage: 12.23x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 25718: contig of 25718 bp in length
* 25719 25818: gap of 100 bp
* 25819 29567: contig of 3749 bp in length
* 29568 29667: gap of 100 bp
* 29668 40520: contig of 10853 bp in length
* 40521 40620: gap of 100 bp
* 40621 50002: contig of 9382 bp in length
* 50003 50102: gap of 100 bp
* 50103 54942: contig of 4840 bp in length
* 54943 55042: gap of 100 bp
* 55043 59067: contig of 4025 bp in length
* 59068 59167: gap of 100 bp
* 59168 72327: contig of 13160 bp in length
* 72328 72427: gap of 100 bp
* 72428 79396: contig of 6969 bp in length
* 79397 79496: gap of 100 bp
* 79497 82738: contig of 3242 bp in length
* 82739 82838: gap of 100 bp
* 82839 87543: contig of 4705 bp in length
* 87544 87643: gap of 100 bp
* 87644 89847: contig of 2204 bp in length
* 89848 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7323 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length
* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2383 bp in length
* 121870 121969: gap of 100 bp
* 121970 125546: contig of 3577 bp in length
* 125547 125646: gap of 100 bp
* 125647 128990: contig of 3344 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length
* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2198 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length


```

* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.
FEATURES
    source
        1. .267581
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone="RP11-133022"
            /clone_lib="RPC1-II.1"
        1. .25718
            /note="assembly_fragment:00645"
            fragment_chain:1
            clone_end:T7
            vector_side:left"
        25819. .29567
            /note="assembly_fragment:01660"
            fragment_chain:1"
        29668. .40520
            /note="assembly_fragment:03080"
            fragment_chain:1"
        40621. .50002
            /note="assembly_fragment:00771"
            fragment_chain:1"
        50103. .54942
            /note="assembly_fragment:02629"
            fragment_chain:1"
        55043. .59067
            /note="assembly_fragment:04565"
            fragment_chain:1"
        59168. .72327
            /note="assembly_fragment:00223"
            fragment_chain:1"
        72428. .79396
            /note="assembly_fragment:03318"
            fragment_chain:1"
        79497. .82738
            /note="assembly_fragment:00224"
            fragment_chain:2"
        82839. .87543
            /note="assembly_fragment:03050"
            fragment_chain:2"
        87644. .89847
            /note="assembly_fragment:01948"
            fragment_chain:3"
        89948. .92376
            /note="assembly_fragment:02191"
            fragment_chain:3"
        92477. .99799
            /note="assembly_fragment:00010"
            fragment_chain:3"
        99900. .102117
            /note="assembly_fragment:00672"
            fragment_chain:3"
        102218. .109330
            /note="assembly_fragment:01326"
            fragment_chain:3"
        109431. .112187
            /note="assembly_fragment:01498"
            fragment_chain:3"
        112288. .116407
            /note="assembly_fragment:01607"
            fragment_chain:3"
        116508. .119386
            /note="assembly_fragment:01899"
            fragment_chain:3"

```

```

misc_feature 119487. .121869
              /note="assembly_fragment:02068"
misc_feature 121970. .125546
              /note="assembly_fragment:02469"
misc_feature 125647. .128990
              /note="assembly_fragment:02502"
misc_feature 129091. .134055
              /note="assembly_fragment:02669"
misc_feature 134156. .138314
              /note="assembly_fragment:02757"
misc_feature 138415. .140612
              /note="assembly_fragment:02776"
misc_feature 140713. .143296
              /note="assembly_fragment:03127"
misc_feature 143397. .148113
              /note="assembly_fragment:03634"
misc_feature 148214. .151009
              /note="assembly_fragment:03898"
misc_feature 151110. .156616
              /note="assembly_fragment:04220"
misc_feature 156717. .158763
              /note="assembly_fragment:04276"
misc_feature 158864. .162129
              /note="assembly_fragment:04418"
misc_feature 162230. .164443
              /note="assembly_fragment:04428"

Query Match 12.2%; Score 100.2; DB 80; Length 267581;
Best Local Similarity 97.1%; Pred. No. 3.3e-17;
Matches 102; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 agaagctgcgcacagaaaacagaggagattgtgtgctgcagcagagagagaccag 60
    |||||
Db 253275 AGAAGCTGCATCGTGGGAAGGACCTGATGATACAGAGGCTCTGTA 253379
    |||||

Qy 61 gaagatctgcattgtgggaagacctgatgatacacagaggtgagaa 105
    |||||
Db 253335 GAAGATCTGCATCGTGGGAAGGACCTGATGATACAGAGGCTCTGTA 253379
    |||||

RESULT 9
AC009556/c 172298 bp DNA HTG 26-MAY-2000
LOCUS Homo sapiens clone RP11-57C21, WORKING DRAFT SEQUENCE, 26 unordered
DEFINITION pieces
ACCESSION AC009556
VERSION AC009556.4 GI:7107752
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 172298)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
TITLE Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
JOURNAL Castle,A., Cerny,J., Collangelo,M., Collins,S., Collimore,A.,
REFERENCE Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo.A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
            Direct Submission
TITLE

```

JOURNAL

COMMENT

Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 28, 2000 this sequence version replaced gi:6479158.
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1672

Center clone name: 57_C_21

----- Summary Statistics

Sequencing vector: M13: M77815; 100% of reads

Chemistry: Dye-terminator-amersham; 5% of reads

Assembly: Dye-terminator Big Dye; 95% of reads

Consensus program: Phrap; version 0.960731

Consensus quality: 136385 bases at least Q40

Consensus quality: 156543 bases at least Q30

Consensus quality: 165534 bases at least Q20

Insert size: 165000; agarose-ff

Insert size: 169798; sum-of-contigs

Quality coverage: 4.2 in Q20 bases; agarose-ff

Quality coverage: 4.1 in Q20 base.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 26 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1016: contig of 1016 bp in length
 * 1017 1116: gap of 100 bp
 * 1117 2136: contig of 1020 bp in length
 * 2137 2236: gap of 100 bp
 * 2237 3571: contig of 1335 bp in length
 * 3572 3671: gap of 100 bp
 * 3672 5272: contig of 1601 bp in length
 * 5273 5372: gap of 100 bp
 * 5373 6381: contig of 1009 bp in length
 * 6382 6481: gap of 100 bp
 * 6482 8508: contig of 2027 bp in length
 * 8509 8608: gap of 100 bp
 * 8609 9886: contig of 1278 bp in length
 * 9887 9986: gap of 100 bp
 * 9987 11943: contig of 1957 bp in length
 * 11944 12043: gap of 100 bp
 * 12044 15112: contig of 3069 bp in length
 * 15113 15212: gap of 100 bp
 * 15213 19323: contig of 4111 bp in length
 * 19324 19423: gap of 100 bp
 * 19424 24065: contig of 4642 bp in length
 * 24066 24165: gap of 100 bp
 * 24166 31658: contig of 7493 bp in length
 * 31659 31758: gap of 100 bp
 * 31759 38737: contig of 6979 bp in length
 * 38738 38837: gap of 100 bp
 * 38838 46131: contig of 7294 bp in length
 * 46132 46231: gap of 100 bp
 * 46232 52344: contig of 6113 bp in length
 * 52345 52444: gap of 100 bp
 * 52445 59763: contig of 7319 bp in length
 * 59764 59863: gap of 100 bp
 * 59864 66816: contig of 6953 bp in length
 * 66817 66916: gap of 100 bp
 * 66917 74010: contig of 7094 bp in length
 * 74011 74110: gap of 100 bp
 * 74111 80599: contig of 6489 bp in length
 * 80600 80699: gap of 100 bp
 * 80700 89507: contig of 8808 bp in length

* 89508 89607: gap of 100 bp
 * 89608 100337: contig of 10730 bp in length
 * 100338 100437: gap of 100 bp
 * 100438 114298: contig of 13861 bp in length
 * 114299 114398: gap of 100 bp
 * 114399 128530: contig of 14132 bp in length
 * 128531 128630: gap of 100 bp
 * 128631 142917: contig of 14287 bp in length
 * 142918 143017: gap of 100 bp
 * 143018 156243: contig of 13226 bp in length
 * 156244 156343: gap of 100 bp
 * 156344 172298: contig of 15955 bp in length.
 * Location/Qualifiers
 * 1..172298
 * /organism="Homo sapiens"
 * /db_xref="taxon:9606"
 * /clone="RP11-57C21"
 * /clone_lib="RPC1-11 Human Male BAC"
 * 1..1016
 * /note="assembly_fragment"
 * 1117..2136
 * /note="assembly_fragment"
 * 2237..3571
 * /note="assembly_fragment"
 * 3672..5272
 * /note="assembly_fragment"
 * 5373..6381
 * /note="assembly_fragment"
 * 6482..8508
 * /note="assembly_fragment"
 * 8609..9886
 * /note="assembly_fragment"
 * 9987..11943
 * /note="assembly_fragment"
 * 12044..15112
 * /note="assembly_fragment"
 * 15213..19323
 * /note="assembly_fragment"
 * 19424..24065
 * /note="assembly_fragment"
 * 24166..31658
 * /note="assembly_fragment"
 * 31759..38737
 * /note="assembly_fragment"
 * 38838..46131
 * /note="assembly_fragment"
 * 46232..52344
 * /note="assembly_fragment"
 * 52445..59763
 * /note="assembly_fragment"
 * 59864..66816
 * /note="assembly_fragment"
 * 66917..74010
 * /note="assembly_fragment"
 * clone_end:T7
 * vector_side:left
 * 74111..80599
 * /note="assembly_fragment"
 * clone_end:SP6
 * vector_side:left
 * 80700..89507
 * /note="assembly_fragment"
 * 89608..100337
 * /note="assembly_fragment"
 * 100438..114298
 * /note="assembly_fragment"
 * 114399..128530
 * /note="assembly_fragment"
 * 128631..142917
 * /note="assembly_fragment"
 * 143018..156243
 * /note="assembly_fragment"
 * 156344..172298
 * /note="assembly_fragment"

FEATURES

source

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

BASE COUNT	51330 a	34914 c	34487 g	49085 t	2502 others
ORIGIN	/note="assembly_fragment"				
Query Match	12.0%;	Score 98.6;	DB 61;	Length 172298;	
Best Local Similarity	96.2%;	Pred. No. 8.8e-17;			
Matches	101;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	1	agaagctggcatcagaaaaacagagggagattgtgtgctgcagccgagggagaccag	60		
Db	16539	AGAAGCTGGCATCAGAAAAACAGAGGGAGATTGTGGCTGCAGCCGAGGAGACCAG	16480		
QY	61	gaagatctgatgtgtggaagacctgatgatcacagaggtgagaa	105		
Db	16479	GAAGATTGCTGTTGGTGAAGACCTGATGATACAGAGGTCTGTA	16435		
RESULT 10					
AF279290	580 bp	DNA	PRI	04-DEC-2000	
LOCUS	Homo sapiens prostate-cancer-specific DD3 protein gene, promoter region and exon 1.				
DEFINITION	AF279290				
ACCESSION	AF279290				
VERSION	AF279290.1 GI:11528086				
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 580)				
AUTHORS	Verhaegh,G.W., van Bokhoven,A., Smit,F., Schaiken,J.A. and Bussemakers,M.J.G.				
TITLE	Isolation and Characterization of the Promoter of the Human Prostate Cancer-specific DD3 Gene				
JOURNAL	J. Biol. Chem. 275 (48), 37496-37503 (2000)				
PUBMED	10982808				
REFERENCE	2 (bases 1 to 580)				
AUTHORS	Verhaegh,G.W., van Bokhoven,A., Smit,F., Schaiken,J.A. and Bussemakers,M.J.G.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-JUN-2000) Urology Research Laboratory, University Medical Center Nijmegen, Geert Grooteplein Zuid 10, Nijmegen 6525 GA, The Netherlands				
FEATURES	Location/Qualifiers				
source	1..580				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="9"				
	/map="9q21-q22"				
promoter	1..460				
protein_bind	260..288				
	/bound_moiety="high mobility group I Y protein"				
mRNA	461..>580				
exon	461..580				
	/number=1				
BASE COUNT	190 a	97 c	153 g	140 t	
ORIGIN					
Query Match	12.0%;	Score 98;	DB 89;	Length 580;	
Best Local Similarity	100.0%;	Pred. No. 5.5e-17;			
Matches	98;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	agaagctggcatcagaaaaacagagggagattgtgtgctgcagccgagggagaccag	60		
Db	483	AGAAGCTGGCATCAGAAAAACAGAGGGAGATTGTGTGGCTGCAGCCGAGGAGACCAG	542		
QY	61	gaagatctgatgtgtggaagacctgatgatcacagag	98		
Db	543	GAAGATCTGCTGTTGGTGAAGACCTGATGATACAGAG	580		

Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B., Brown, A., Castle, A., Collangelo, M., Collins, S., Collimore, A., Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 22, 2000 this sequence version replaced gi:7259702.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE JOURNAL COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L1240
Center clone name: 179_B-14
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 168987 bases at least Q40
Consensus quality: 179731 bases at least Q30
Consensus quality: 185225 bases at least Q20
Insert size: 183000; agarose-tp
Insert size: 189119; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-tp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
13 112: contig of 12 bp in length
13 112: gap of 100 bp
113 1767: contig of 1655 bp in length
1768 1867: gap of 100 bp
1868 3764: contig of 1897 bp in length
3765 3864: gap of 100 bp
3865 6037: contig of 2173 bp in length
6038 6137: gap of 100 bp
6138 8137: contig of 2000 bp in length
8138 8237: gap of 100 bp
8238 10419: contig of 2182 bp in length
10420 10519: gap of 100 bp
10520 13978: contig of 3459 bp in length
13979 14078: gap of 100 bp
14079 41398: contig of 27320 bp in length
41399 41498: gap of 100 bp
41499 45869: contig of 4371 bp in length
45870 45969: gap of 100 bp
45970 49945: contig of 3976 bp in length
49946 50045: gap of 100 bp
50046 56380: contig of 6335 bp in length
56381 56480: gap of 100 bp
56481 60955: contig of 4475 bp in length
60956 61055: gap of 100 bp
61056 66082: contig of 5027 bp in length

* 66083 66182: gap of 100 bp
66183 74638: contig of 8456 bp in length
74639 74738: gap of 100 bp
74739 81245: contig of 6507 bp in length
81246 81345: gap of 100 bp
81346 89538: contig of 8193 bp in length
89539 89638: gap of 100 bp
89639 102730: contig of 13092 bp in length
102731 102830: gap of 100 bp
102831 115832: contig of 13002 bp in length
115833 115932: gap of 100 bp
115933 129142: contig of 13210 bp in length
129143 129242: gap of 100 bp
129243 146224: contig of 16982 bp in length
146225 146324: gap of 100 bp
146325 167164: contig of 20840 bp in length
167165 167264: gap of 100 bp
167265 188461: contig of 21197 bp in length
188462 188561: gap of 100 bp
188562 191319: contig of 2758 bp in length.

FEATURES source

1. 191319
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-179B14"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature
1. 12
/note="assembly_fragment"
clone_end:SP6
vector_side:left
misc_feature
113. 1767
/note="assembly_fragment"
misc_feature
1868. 3764
/note="assembly_fragment"
misc_feature
3865. 6037
/note="assembly_fragment"
misc_feature
6138. 8137
/note="assembly_fragment"
misc_feature
8238. 10419
/note="assembly_fragment"
misc_feature
10520. 13978
/note="assembly_fragment"
misc_feature
14079. 41398
/note="assembly_fragment"
misc_feature
41499. 45869
/note="assembly_fragment"
misc_feature
45970. 49945
/note="assembly_fragment"
misc_feature
50046. 56380
/note="assembly_fragment"
misc_feature
56481. 60955
/note="assembly_fragment"
misc_feature
61056. 66082
/note="assembly_fragment"
misc_feature
66183. 74638
/note="assembly_fragment"
misc_feature
74739. 81245
/note="assembly_fragment"
misc_feature
81346. 89538
/note="assembly_fragment"
misc_feature
89639. 102730
/note="assembly_fragment"
misc_feature
102831. 115832
/note="assembly_fragment"
misc_feature
115933. 129142
/note="assembly_fragment"
misc_feature
129243. 146224
/note="assembly_fragment"
misc_feature
146325. 167164
/note="assembly_fragment"
misc_feature
167265. 188461

THIS PAGE BLANK (USPTO,

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 04:11:52 ; Search time 6114.61 Seconds
(without alignments)
5537.577 Million cell updates/sec

Title: US-09-402-713A-6
Perfect score: 3582
Sequence: 1 acagaagaataagcaagtgc.....tgattctttgttacaacttt 3582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_estl:*

2: gb_est2:*

3: gb_est3:*

4: gb_est4:*

5: gb_est5:*

6: gb_est6:*

7: gb_est7:*

8: gb_est8:*

9: gb_est9:*

10: gb_est10:*

11: gb_est11:*

12: gb_est12:*

13: gb_est13:*

14: gb_est14:*

15: gb_est15:*

16: gb_est16:*

17: gb_est17:*

18: gb_est18:*

19: gb_est19:*

20: gb_est20:*

21: gb_est21:*

22: gb_est22:*

23: gb_est23:*

24: gb_est24:*

25: gb_est33:*

26: gb_est34:*

27: gb_est35:*

28: gb_est36:*

29: gb_est37:*

30: gb_est38:*

31: gb_est39:*

32: gb_est40:*

33: em_estba:*

34: em_estfun:*

35: em_esthum1:*

36: em_esthum2:*

37: em_esthum3:*

38: em_esthum4:*

39: em_esthum5:*

40: em_esthum6:*

41: em_esthum7:*

42: em_esthum8:*

43: em_esthum9:*

44: em_esthum10:*

45: em_esthum11:*

46: em_esthum12:*

47: em_esthum13:*

48: em_esthum14:*

49: em_esthum15:*

50: em_esthum16:*

51: em_esthum17:*

52: em_esthum18:*

53: em_esthum19:*

54: em_esthum20:*

55: em_esthum21:*

56: em_esthum22:*

57: em_esthum23:*

58: em_esthum24:*

59: em_esthum25:*

60: em_esthum26:*

61: em_esthum27:*

62: em_esthum28:*

63: em_estin1:*

64: em_estin2:*

65: em_estin3:*

66: em_estin4:*

67: em_estin5:*

68: em_estomi:*

69: em_estom2:*

70: em_estov1:*

71: em_estov2:*

72: em_estpl1:*

73: em_estpl2:*

74: em_estpl3:*

75: em_estpl4:*

76: em_estpl5:*

77: em_estpl6:*

78: em_estpl7:*

79: em_estpl8:*

80: em_estpl9:*

81: em_estpl10:*

82: em_estro1:*

83: em_estro2:*

84: em_estro3:*

85: em_estro4:*

86: em_estro5:*

87: em_estro6:*

88: em_estro7:*

89: em_estro8:*

90: em_estro9:*

91: em_estro10:*

92: em_estro11:*

93: em_estro12:*

94: em_estro13:*

95: em_estro14:*

96: em_estro15:*

97: em_estro16:*

98: em_estro17:*

99: em_estro18:*

100: em_estro19:*

101: em_estro20:*

102: gb_est25:*

103: gb_est26:*

104: gb_est27:*

105: gb_est28:*

106: gb_est29:*

107: gb_est30:*

108: gb_est31:*

109: gb_est32:*

110: gb_est41:*

111: gb_est42:*

112: gb_est43:*

113: gb_est44:*

114: gb_est45:*

115: gb_est46:*

116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

COMMENT

Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@igr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	615.8	17.2	654	227	AQ319247	AQ319247 RPCI11-10
c 2	536	15.0	546	227	AQ318289	AQ318289 RPCI11-10
c 3	533	14.9	542	227	AQ318297	AQ318297 RPCI11-10
c 4	514.2	14.4	856	21	AI557225	AI557225 PT2_1_15
c 5	401.2	11.2	441	170	BF858890	BF858890 RC5-FT019
c 6	394	11.0	402	9	AA578773	AA578773 nh24a04.s
c 7	362	10.1	394	225	AQ206972	AQ206972 HS_3238_B
c 8	338.4	9.4	346	122	AW948945	AW948945 QV4-FT000
c 9	331	9.2	451	226	AQ246715	AQ246715 HS_2059_B
c 10	306.2	8.5	332	170	BF858286	BF858286 RC5-FT019
c 11	260.4	7.3	290	147	BF373619	BF373619 MR0-FT017
c 12	226.8	6.3	282	147	BF373581	BF373581 MR0-FT017
c 13	214.4	6.0	226	170	BF858371	BF858371 RC5-FT019
c 14	167	4.7	167	147	BF373406	BF373406 IL2-FT015
c 15	160.2	4.5	630	226	AQ302255	AQ302255 HS_3141_B
c 16	86.4	2.4	657	21	AI557495	AI557495 PT2_1_7_H
c 17	63.2	1.8	1101	219	CNS00396	AL063921 Drosophil
c 18	62.6	1.7	600	232	AQ673276	AQ673276 HS_5469_B
c 19	61	1.7	376	223	AQ069980	AQ069980 HS_3027_B
c 20	59.6	1.7	1067	219	CNS00C08	AL058564 Drosophil
c 21	58.8	1.6	997	219	CNS005TE	AL060767 Drosophil
c 22	56.8	1.6	1001	219	CNS01400	AL103554 Drosophil
c 23	56.2	1.6	1101	219	CNS000D1	AL085414 Drosophil
c 24	55	1.5	463	257	B81447	B81447 CIT-HSP-205
c 25	52.8	1.5	1101	219	CNS0100X	AL098379 Drosophil
c 26	52.6	1.5	1101	219	CNS0182P	AL108811 Drosophil
c 27	51.4	1.4	987	219	CNS00418	AL066537 Drosophil
c 28	51	1.4	1034	220	CNS02N17	AL205432 Tetradon
c 29	50.2	1.4	1101	219	CNS0039E	AL063919 Drosophil
c 30	49.6	1.4	958	221	CNS03CMC	AL237981 Tetradon
c 31	49.2	1.4	1101	219	CNS0105X	AL098595 Drosophil
c 32	49	1.4	1101	219	CNS017KX	AL108171 Drosophil
c 33	49	1.4	1204	219	CNS016E2	AL106628 Drosophil
c 34	48.6	1.4	1090	221	CNS04U88	AL307457 Tetradon
c 35	48.4	1.4	1086	192	AQ020508	Mus muscu
c 36	47.6	1.3	529	225	AQ173559	AQ173559 HS_3202_A
c 37	47.4	1.3	1458	192	AQ018211	Mus muscu
c 38	47.2	1.3	586	235	AQ928836	RPCI-23-2
c 39	47.2	1.3	659	242	AZ358277	1M0100F09
c 40	47.2	1.3	677	242	AZ357710	1M0099C05
c 41	46.8	1.3	654	233	AQ046642	RPCI11-35
c 42	46.8	1.3	713	240	AZ253521	RPCI-23-4
c 43	46.8	1.3	854	140	BE777727	601348546
c 44	46.8	1.3	1101	219	CNS00067	AL062049 Drosophil
c 45	46.6	1.3	421	163	BE137644	u966f11.y

ALIGNMENTS

RESULT 1	AQ319247/c	AQ319247	654 bp	DNA	GSS	06-MAY-1999
LOCUS	RPCI11-108L4.TV	RPCI-11	Homo sapiens	genomic clone	RPCI-11-108L4,	
DEFINITION	DNA sequence.					
ACCESSION	AQ319247					
VERSION	AQ319247.1	GI:4052212				
KEYWORDS	GSS.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 654)					
TITLE	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.					
JOURNAL	Use of human BAC End Sequences for Sequence-Ready Map Building Unpublished (1998)					

```
QY 1816 tatacttcattctctctctctctctcaaatatccaaacgaagcttttcacagaattc 1870
|||||
Db 56 TATACTTCATTCTCTATCTCTATCAATAATCAACAAGCTTTTCAGAGAAATC 2

RESULT 2
AQ318289/c
LOCUS AQ318289 546 bp DNA GSS 04-MAY-1999
DEFINITION RPCI11-108A18.TV RPCI-11 Homo sapiens genomic clone RPCI-11-108A18,
DNA sequence.
ACCESSION AQ318289
VERSION AQ318289.1 GI:4050259
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 546)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: RPCI11-108A18.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..546
/organism="Homo sapiens"
/db_xref="GDB:7541105"
/db_xref="taxon:9606"
/clone="RPCI-11-108A18"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 150 a 106 c 116 g 172 t 2 others
ORIGIN

Query Match 15.0%; Score 536; DB 227; Length 546;
Best Local Similarity 98.7%; Pred. No. 3.2e-130;
Matches 539; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2287 aaagcagctggaatggacacacaaatgcataatcgtatactctaccatcagtcata 2346
|||||
Db 546 AAAGCAGCTGGAATGGACACACAAATATGCATAATCTCAATCTTACCATCAGCTACA 487
|||||
QY 2347 cactgcttgacatatattgttagaagcacctcgcatttgggttctcttaagcaaaata 2406
|||||
Db 486 CACTGCTTGACATATATTGTAGAACACCTCGCAITTTGGGGTCTCTTTAGCAAAATA 427
|||||
QY 2407 cttgcattaggctcagctggggctgtgcatcagcggtttggagaataattcaattctca 2466
|||||
Db 426 CTTGCAATTAGGCTCAGCTGGGGCTGTGCATCAGNCGGTTTGAGAAATATTCAATTCTCA 367
|||||
QY 2467 gcagaagccagaatttgaattccctcatcttttagaatacattaccaggtttggagagg 2526
|||||
Db 366 GCAGAAGCCAGAATTTGAATTCCTCATCTTTTAGGAATCATTTACCAGGTTGGAGAG 307
```

```
QY 2527 attcagacagctcaggtgttttcaactaagtctctgaacttctgtccctcttggttca 2586
|||||
Db 306 ATTCAAGACAGCTCAGGTGCTTTTCACATAATGTCTCTGAACCTTCTGTCCTCTTTGTGTTCA 247
|||||
QY 2587 tggatagtcocataataatgtttattcttttgaactgatgctcatagagagaataaagaa 2646
|||||
Db 246 TGGATAGTCCAATAAATAATGTATTCTTTGAAGCTGATGCTCATAGGAGAGAAATAAGAA 187
|||||
QY 2647 ctctgagtcatatcaacattaggattcaagaataattagatttaagctcacactggctc 2706
|||||
Db 186 CTCGTAGTGATATCAACATTAGGGATTCAAGAAATATTAGATTAAAGCTCACACGTGGTC 127
|||||
QY 2707 aaaggaaccaagatacacaagaactctgagctgctatcgctcccatctctgtgagccaca 2766
|||||
Db 126 AAAAGGAACCAAGATACAAAGAACTCTGAGCTGCTCATGTCCTCCCATCTCTGTGAGCCACA 67
|||||
QY 2767 accaacagcaggagcccaacgcagtctgagatccttaaatcaaggaaacacagtgctcatga 2826
|||||
Db 66 ACCAACAGCAGGAGCCCAACGATGCTGTGAGATCTTTAAATCAAGGAAACCAAGTGTCTATGA 7
|||||
QY 2827 gttgaa 2832
|||||
Db 6 GTTGAA 1

RESULT 3
AQ318297/c
LOCUS AQ318297 542 bp DNA GSS 04-MAY-1999
DEFINITION RPCI11-108A20.TV RPCI-11 Homo sapiens genomic clone RPCI-11-108A20,
DNA sequence.
ACCESSION AQ318297
VERSION AQ318297.1 GI:4050267
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 542)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: RPCI11-108A20.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..542
/organism="Homo sapiens"
/db_xref="GDB:7541107"
/db_xref="taxon:9606"
/clone="RPCI-11-108A20"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 151 a 106 c 118 g 167 t
ORIGIN
```

Query Match 14.9%; Score 533; DB 227; Length 542;
Best Local Similarity 100.0%; Pred. No. 2e-129;
Matches 533; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2303 gacacacacataatgcataaaactaacctccaccatcagctacacacacgcttgacatata 2362
DB 542 GACAAACCAATATATCAAAATCTAACTCCCTACCATAGCTACACACTGCTTGACATATA 483
QY 2363 ttgttagagacacctcgatttgggttctcttaagcaaaatcttcattagctccta 2422
DB 482 TTGTTAGAGACACCTCGCATTTGCTGGTTCCTTAAAGCAAAATATCTTGCATTAGTCTCA 423
QY 2423 gctgggggtgtgcacagcggtttgagaataattcttcagcagaagcagaattt 2482
DB 422 GCTGGGGCTGTGCATCAGCGGCTTTGAGAAATATTTCAATTCACAGAAAGCCAGAAATTT 363
QY 2483 gaattccctcatcttttaggaataattaccaggtttgagagagattcagacagctcagg 2542
DB 362 GAATTCCTCTCATCTTTTAAAGGAATCAATTTACCAGGTTTGAGAGAGATTGAGACAGCTCAGG 303
QY 2543 tgccttactaatgtctctgaactctgtccctcttctgttctcatggtatgccaataaa 2602
DB 302 TGCTTTCACTAATGTCTCTGAACCTTCTGTCCCTCTTGTGTTTCATGGATAGTCCCAATAA 243
QY 2603 taatgttatctttgaactgactgcctcatagggagagagaataagaacctctgagtgatataca 2662
DB 242 TAATGTTATCTTTGAACCTGATCTCATAGGAGAGAATAAAGAACTCTGAGTGATATCAA 183
QY 2663 cattaggattcaagaagaataattagatttaagctcacactgctgcaaaaggaaacaaagata 2722
DB 182 CATTTAGGAGATCAAAAGAAATATTAGATTAAAGCTCACACTGGTCAAAAGGAACCAAGATA 123
QY 2723 caaagaactctgagctgcctcctcatctctgtgagccacaacacagcagacc 2782
DB 122 CAAGAAGCTTGAGCTGTCATCGTCCCATCTCTGTGAGCCCAACCAACAGCAGACCC 63
QY 2783 aacgcatctctgagatccttaatacaagaaacacagtgctcatgagtgaaattc 2835
DB 62 AACGCATGCTGAGATCCTTAATCAAGGAACCAAGTGTGATGAGTGGAATTC 10

RESULT 4
AI557225 856 bp mRNA EST 09-AUG-1999
LOCUS PT2.L15_B05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
DEFINITION AI557225
ACCESSION AI557225
VERSION AI557225.1 GI:4489588
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 856)
AUTHORS Huang,G.M., Ng,W.I., Parks,J., He,L., Liang,H.A., Gordon,D., Yu,J.
and Hood,L.
TITLE Prostate cancer expression profiling by cDNA sequencing analysis
JOURNAL Genomics 59 (2), 178-186 (1999)
MEDLINE 99339982
COMMENT Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.
Location/Qualifiers
1. .856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional source

cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."

BASE COUNT 237 a 191 c 180 g 214 t 34 others
ORIGIN

Query Match 14.4%; Score 514.2; DB 21; Length 856;
Best Local Similarity 97.1%; Pred. No. 2.1e-124;
Matches 575; Conservative 0; Mismatches 11; Indels 6; Gaps 5;

QY 1352 actaagtcctttatccctcccttctgttggattttttccagtgataaaagttaaaatgct 1411
DB 18 ACGTAGCCCTTAAATCCCTCCCTCTTTGATTTTTCAGTATATAAAAGTTAAAAATGCT 77
QY 1412 tagccttgactgagcgtgtatcacag-cacagcctctcccccctccctccagctctatcg 1470
DB 78 TAGCCTTGTACTGAGCTGTATACAGCCACAGCCTCTCCCATCCCTCCAGCCTTATCTG 137
QY 1471 tcatcaccatcaacccctcccatcaccacatacaaaaaatccttaacttgccttgaac 1530
DB 138 TCATCACCATCAACCCCTCCCAT-GCACCTAAACAAAATCTAACCTGTAATTCCTTTGAAC 196
QY 1531 atgtcaggacatacatattctcttctgctgagagctcttctctgtctcttaaatctag 1590
DB 197 ATGTCAGG-CATACATTAATTCCTCTGCTGAGAAAGCTCTCTCTGCTCTTTAAATCTAG 255
QY 1591 aatgatgtaaagtttgaataaagtgtgactatcttcttcatacaagaagggacacatat 1650
DB 256 AATGATGTAAAGTTTGAATAAGTTGACTATCTTACTTCATGCAAGAAGGGACACATAT 315
QY 1651 gagattcatcatcatgagagagacaaatactaaagtgttaattgtattataagagttta 1710
DB 316 GAGATTCATCATCATGAGAGACAGCAAAATCTAAAGTGTAAATTTGATTATTAAGAGTTTA 375
QY 1711 gataatatatgaatgcaagagccacagagggaaatgtttatggygcagctttgtaagcc 1770
DB 376 GATAAATATATGAAATGCAAGAGCCACAGAGGAATGTTTATGGGCACTGTTTGAAGCC 435
QY 1771 tgggatgtgaagcaagcaggaacccctcatgacttcttataataatacttctcttc 1830
DB 436 TGGGATGTGAAGCAAGGCGGACCTCATAGTATCTATATAATATATCTTCAATCTC 495
QY 1831 tatctctatcaaatatccaaagcttttccagaatctcatgagtcagtcgcaatcccca 1890
DB 496 NATCTCTATCAATATATCAACAAAGCTTTTCACAGAAATTCATGCAAGTCCCAAA 555
QY 1891 ggtaacctttat-ccatttcattggtgagtcgcttt--agaattttggcaaa 1939
DB 556 GGGAACTTTATCCCATTTTCATGTTGAGTGGCGCTTNAAGAAATTTTGGNAA 607

RESULT 5
BF58890/c 441 bp mRNA EST 16-JAN-2001
LOCUS RC5-FT0194-071200-023-G11 FT0194 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF58890
ACCESSION BF58890
VERSION BF58890.1 GI:12246634
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 441)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Db 241 TTCATACCAAAATCATTTTCATATATTTTAAACCTCAAAACAAAGCTGTGTCTAATATCTGAT 300

Qy 882 ctctacgattctcttgccgcaacattctccatatatccagccacacactcatcttttaata 941

Db 301 CTTCTACGGTTCCTCTGGGCCAACATTTCTCATATATCCAGCCACACTCATTTTAAATA 360

Qy 942 tttagtctccagatctgtactgtgacctttctacactgtaga 983

Db 361 TTTAGTCTCCAGATCTGTACTGTGACCTTTCTACACTGTAGA 402

RESULT 7

LOCUS AQ206972/c 394 bp DNA GSS 17-SEP-1998

DEFINITION HS_3238_B1.G11_MR_CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3238 Col=21 Row=N, DNA sequence.

ACCESSION AQ206972

VERSION AQ206972.1 GI:3617542

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 394)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 3238 Row: N Column: 21 Class: BAC ends High quality sequence stop: 394. Location/Qualifiers 1. 394 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="Plate-3238 Col=21 Row=N" /sex="male" /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 134 a 62 c 93 g 103 t 2 others

ORIGIN source

Query Match 10.1%; Score 362; DB 225; Length 394; Best Local Similarity 97.7%; Pred. No. 2,3e-84; Matches 387; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

Qy 892 cttctgggcccacattctccatatatccagccacactcatcttttaatttagtctcc 951

Db 394 CCGTCTGGGCNACATTTCCATATATCCAGCCACACTCATTTTAAATATTAGTCC 335

Qy 952 agatctgtactgtacctttctacactgtagataacattactcatctttgttccaagacc 1011

Db 334 AGATCTGTACTGTGACCTTCTACACTGTGAGTAATAACTTACTATTGTTCAAAGACC 275

Qy 1012 ctctgtgtgtcgtcctaataatagctactgtttttctcaaggagtgcttgcccaagg 1071

Db 274 CTTCTGTTGTCTGCCTTAATATAGTACTGTCTTTTCAATAGGAGTGTCTGGCCAGG 215

Qy 1072 ggaatctgtgaacaggctgggaagcatctcaagatctcttccagggttatacttactagcac 1131

Db 214 GGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGTTATATCTACTAGCAC 155

Qy 1132 acagcatatcatcagtgaggaattatctaatcaacatcatcctcagtgctcttgccca 1191

Db 154 ACAGCATGATCATTTACGGAGTGAAATTATATCAATCAACATCATCTTCAGTGCTCTTGCCCA 95

Qy 1192 tactgaattcatcttcccacttttgcctatttgcctatttcaagacctcaaaaatgtctaccatt 1251

Db 94 TACTGAATTCATTTCCACATTTTCTGTG-CCATTTCTCAAGA-CTCAAAATGTCTATTCATT 37

Qy 1252 aatcacagaggaattaaacttttttttttaacctggaa 1287

Db 36 AATATCACAGGATTAACCTTTTTTTTTTTTACCCCTGCAA 1

RESULT 8

LOCUS AW948945 346 bp mRNA EST 31-MAY-2000

DEFINITION QV4-FT0004-130500-212-d11 FT0004 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW948945

VERSION AW948945.1 GI:8126719

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 346)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV4-FT0004-130500-212-d11&t3=2000-05-13&t4=1>) Seq primer: puc 18 forward High quality sequence stop: 344. Location/Qualifiers 1. 346 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="FT0004" /dev_stage="Adult" /note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 104 a 81 c 63 g 98 t

ORIGIN source

Query Match 9.4%; Score 338.4; DB 122; Length 346; Best Local Similarity 99.7%; Pred. No. 3.7e-78; Matches 339; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2591 tagtccaataaataatgttattctttgaactgatgcctagaggagagataataagaactct 2650

Db	7	TAGTCCAATAATAATGTTACCTTTGAACTGATGCTCATAGAGAGAATAAAGAACTCT	66
Qy	2651	gagtgatatcaacattagggattcaaaagaataattagatttaaagctcacactggctcaaaa	2710
Db	67	GAGTGATATCAACATTAGGGATTCAAGAATAATTAGATTAAAGCTCACACTGGTCAAAA	126
Qy	2711	ggaaccaagatacaaaagaactctgagctgtcatcgtcccatctctctgtgagccacaacca	2770
Db	127	GGACCAAGATACAAAGAACTCTGAGCTGTCTATCGTCCCATCTCTGTGAGCCACAACCA	186
Qy	2771	acagcagagacccaacgcatgctgagatcccttaataatcaaggaacacagtgcatgagttg	2830
Db	187	ACAGCAGAGACCAACGCAATGCTGAGATPCCCTTAAATCAAGGAAACCAAGTGCATGAGTTG	246
Qy	2831	aattctctattatgtagctgactctctgagcactctctgctctcctcttgacacattt	2890
Db	247	AATCTCTATTATGGATGCTAGCTTTCGGCCATCTCTGGCTCTCCTTTGACACATATT	306
Qy	2891	agctctagccttggctccacgactttatctttctccc	2930
Db	307	AGCTTCTAGCCTTTGCTTCCACGACTTTTATCTTTTCTCC	346
RESULT	9		
AQ246715		451 bp DNA	GSS 06-OCT-1998
LOCUS	HS_2059_B2_E08_T7	CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2059 Col=16 Row=J, DNA sequence.	
DEFINITION	AQ246715		
ACCESSION	AQ246715.1	GI:3696897	
VERSION		GSS.	
KEYWORDS		human.	
SOURCE		Homo sapiens	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE		1 (bases 1 to 451)	
AUTHORS		Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.	
TITLE		Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome	
JOURNAL		Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)	
MEDLINE		99380589	
COMMENT		Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 2059 row: J column: 16 Class: BAC ends High quality sequence stop: 451. Location/Qualifiers 1. .451 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="plate=2059 Col=16 Row=J" /clone_lib="CIT Approved Human Genomic Sperm Library D" /sex="male" /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in, E-Coli DH10B"	
FEATURES			
source			
BASE COUNT	133 a	101 c	86 g 130 t 1 others
ORIGIN			
Query Match		9.2%;	Score 331; DB 226; Length 451;
Best Local Similarity		90.1%;	Pred. No. 3.6e-76;
Matches 411; Conservative		0; Mismatches 36; Indels 9; Gaps 5;	
Qy 1908	tcatggtgagtcgcgttagaatttgggcaaatcatactggtcacttatctcacttga	1967	

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asampson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&st2=MR0&st3=MR0&st4=1>)
210800-101-d05&st3=2000-08-21&st4=1
Seq primer: puc 18 forward
High quality sequence start: 35
High quality sequence stop: 282.

FEATURES

Source	Location/Coordinates
1. .282	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone_lib="FT0175"
	/dev_stage="Adult"
	/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ARESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT	85 a 51 c 70 q 76 t

BASE COUNT	85 a	51 c	70 g	76 t
ORIGIN				
Query Match		6.3%	Score 226.8;	DB 147;
Best Local Similarity		91.3%;	Pred. No. 9.5e-49;	Length 282;
Matches 252;	Conservative	0;	Mismatches 22;	Indels 2;
Gaps 1;				

Qy	755	caagctggggagagataaccacgaggcgagagggtcagagattctggccctgtgctctaa	814
Db	282	CAAGCTGGGGAGGAGATAAACACGGGGCAGAGGGTCAGGATTTGGCCCTGCTGCTAA	223
Qy	815	ctgtcggttcataaccacaatcatttcatattcttaacctcaaaacaaagctgttgtaat	874
Db	222	CTGTGGGTTCATAAACCAATCATTTTCATATTTCTAACCTCAAAACAAAGCTGTGTAT	163
Qy	875	atctgatctctagtggttccttcttctggggcccaacattctctcatatctccagccacactcatt	934
Db	162	ATCTGATCTCTAGGGTTCCTTCTGGGCCCAAGATTTCTCCATATATCCAGCCACACTCAT	103
Qy	935	tttaataattagttccagatctgtactgtgaccttgc--tacactgtagaataacatta	992
Db	102	TTTAAATATTTAGTTCACGATCTGTACTGTGACCTTCTCTACATCTGTGAGAATAACATTA	43
Qy	993	ctcattttgttcaaaagacctctgtgtgtgctgcta	1028
Db	42	CTCATTTGTTTCAAGACCTTTCACCTTCTGCTAGGCACTA	7

RESULT	13
BF858371/c	
LOCUS	226 bp mRNA EST 16-JAN-2001
DEFINITION	RCS-FTO193-211100-012-E11 FTO193 Homo sapiens cDNA, mRNA sequence.
ACCESSION	BF858371
VERSION	BF858371.1 GI:12246115
KEYWORDS	EST.
SOURCE	human.

ORGANISM

Homio sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 226)

Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zagó, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

AUTHORS

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE	COMMENT
107689	107689
107690	107690
107691	107691
107692	107692
107693	107693
107694	107694
107695	107695
107696	107696
107697	107697
107698	107698
107699	107699
107700	107700
107701	107701
107702	107702
107703	107703
107704	107704
107705	107705
107706	107706
107707	107707
107708	107708
107709	107709
107710	107710
107711	107711
107712	107712
107713	107713
107714	107714
107715	107715
107716	107716
107717	107717
107718	107718
107719	107719
107720	107720
107721	107721
107722	107722
107723	107723
107724	107724
107725	107725
107726	107726
107727	107727
107728	107728
107729	107729
107730	107730
107731	107731
107732	107732
107733	107733
107734	107734
107735	107735
107736	107736
107737	107737
107738	107738
107739	107739
107740	107740
107741	107741
107742	107742
107743	107743
107744	107744
107745	107745
107746	107746
107747	107747
107748	107748
107749	107749
107750	107750
107751	107751
107752	107752
107753	107753
107754	107754
107755	107755
107756	107756
107757	107757
107758	107758
107759	107759
107760	107760
107761	107761
107762	107762
107763	107763
107764	107764
107765	107765
107766	107766
107767	107767
107768	107768
107769	107769
107770	107770
107771	107771
107772	107772
107773	107773
107774	107774
107775	107775
107776	107776
107777	107777
107778	107778
107779	107779
107780	107780
107781	107781
107782	107782
107783	107783
107784	107784
107785	107785
107786	107786
107787	107787
107788	107788
107789	107789
107790	107790
107791	107791
107792	107792
107793	107793
107794	107794
107795	107795
107796	107796
107797	107797
107798	107798
107799	107799
107800	107800
107801	107801
107802	107802
107803	107803
107804	107804
107805	107805
107806	107806
107807	107807
107808	107808
107809	107809
107810	107810
107811	107811
107812	107812
107813	107813
107814	107814
107815	107815
107816	107816
107817	107817
107818	107818
107819	107819
107820	107820
107821	107821
107822	107822
107823	107823
107824	107824
107825	107825
107826	107826
107827	107827
107828	107828
107829	107829
107830	107830
107831	107831
107832	107832
107833	107833

2020263
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-FR0193-211100-012-E11&t3=2000-11-21&t4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 226.

FEATURES

```

FEATURES
i. .226
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0193"
/dev_stage="Adult"
/site="Organ; prostate.tumor; Vector; puc18; Site_1; SmaI;
Site_2; SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 136,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
79 a 29 c 59 g 59 t
BASE COUNT
ORIGIN

```

	Query Match	6.08;	Score 214..4;	DB 170;	Length 226;
	Best Local Similarity 99.5%;	Pred. No. 1.7e-45;	Mismatches 0;	Indels 1;	Gaps 0;
	Matches 215; Conservative				
QY	809	cctaaactgtgcgctcataaccaaaatcatttcatttctaacctcaaacacaaagctgt	868		
Db	226	CCTAAACTGTCGGTTCATACCAAATCATTTTCATTTCATAACCCTCANAACAAGCTGT	167		
QY	869	tgtaatactgatctctcacggtttcctcttgaggccacaacattctccatatatcccaggcca	928		
Db	166	TGTAATAATCATGCCTCACCGTTCCTCTTGCGGCCAACATTCCTCATATATCCAGCCACA	107		
QY	929	ctcaattttaaatattagtctccagatctgaactgtgaacttctcacactgagaataac	988		
Db	106	CTCATTTTTTAATATTTAGTTCCCAGATCTGTACTGTGACCTTCTTACACTGTAGAATAAC	47		
QY	989	attaactatttgttcaaagacccttcggtgtggtg	1024		
Db	46	ATTACTCATTTTGTTCAAAGACCCCTCGGTGGCAG	11		

RESULT	14
BF373406	
LOCUS	BF373406
DEFINITION	IL2-FT0159-070800-120-H01 FT0159 Homo sapiens CDNA, mRNA sequence.
ACCESSION	BF373406
VERSION	BF373406.1
KEYWORDS	EST.
SOURCE	human.
	167 bp mRNA EST 24-NOV-2000

SEQUENCE ORGANISM	REFERENCE AUTHORS	TITLE
Homo sapiens		Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Theria; Primates; Catarrhini; Homnidae; Homo.		
1 (bases 1 to 167)		
Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F.F., de Souza,S.J. and Simpson,A.J.		

THIS PAGE BLANK (USP)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:25:42 ; Search time 176.8 Seconds
(without alignments)
3835.478 Million cell updates/sec

Title: US-09-402-713A-6
Perfect score: 3582
Sequence: 1 acgaagaataagcaagtc.....tgattctttgttacaacttt 3582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	77.4	2.2	7218	1 US-08-232-463-14	Sequence 14, Appl
2	57.4	1.6	7218	1 US-08-232-463-14	Sequence 14, Appl
3	43	1.2	10968	2 US-08-680-327-2	Sequence 2, Appl
4	43	1.2	10968	4 US-09-228-246-1	Sequence 1, Appl
5	41.6	1.2	10409	3 US-08-772-440-33	Sequence 33, Appl
6	41.4	1.2	5475	2 US-08-680-327-1	Sequence 1, Appl
7	41.4	1.2	5475	4 US-09-228-246-3	Sequence 3, Appl
8	40.6	1.1	278	1 US-08-222-177A-52	Sequence 52, Appl
9	40.6	1.1	3757	2 US-09-016-366A-13	Sequence 13, Appl
10	40.6	1.1	3757	2 US-08-978-404B-19	Sequence 19, Appl
11	40.6	1.1	43795	3 US-08-742-185-101	Sequence 101, Appl
12	40.4	1.1	5134	2 US-08-310-912A-157	Sequence 157, Appl
13	40.4	1.1	5134	5 PCT-US95-04589-157	Sequence 157, Appl
14	40	1.1	22846	2 US-08-469-461-3	Sequence 3, Appl
15	40	1.1	22846	3 US-07-890-609-3	Sequence 3, Appl
16	38.8	1.1	177	2 US-08-829-961-3	Sequence 3, Appl
17	38.8	1.1	2618	4 US-09-488-671-17	Sequence 17, Appl
18	38.8	1.1	12141	4 US-09-488-671-10	Sequence 10, Appl
19	38	1.1	26700	1 US-08-472-217-1	Sequence 1, Appl
20	38	1.1	26700	2 US-08-488-199-5	Sequence 5, Appl
21	38	1.1	26700	3 US-08-760-534A-1	Sequence 1, Appl
22	37.6	1.0	2529	3 US-09-051-969A-5	Sequence 5, Appl
23	37.6	1.0	11236	1 US-07-853-913-1	Sequence 1, Appl
24	37.4	1.0	2775	1 US-08-149-096A-1	Sequence 1, Appl
25	37.2	1.0	223	1 US-08-222-177A-14	Sequence 14, Appl
26	37.2	1.0	3588	1 US-08-197-792-32	Sequence 32, Appl
27	37.2	1.0	3588	1 US-08-459-850-32	Sequence 32, Appl

c 28	37.2	1.0	3588	1	US-08-459-214-32	Sequence 32, Appl
c 29	37	1.0	265	1	US-08-222-177A-1	Sequence 1, Appl
c 30	37	1.0	1493	1	US-08-340-820-24	Sequence 24, Appl
c 31	37	1.0	1493	1	US-08-593-535-24	Sequence 24, Appl
c 32	36.6	1.0	6350	2	US-08-385-335A-9	Sequence 9, Appl
c 33	36.4	1.0	298	1	US-08-599-252-88	Sequence 88, Appl
c 34	36.4	1.0	298	5	PCT-US96-06352-88	Sequence 88, Appl
c 35	36.4	1.0	298	5	PCT-US96-06352-88	Sequence 88, Appl
c 36	36.4	1.0	600	1	US-08-599-252-104	Sequence 104, Appl
c 37	36.4	1.0	600	5	PCT-US96-06352-104	Sequence 104, Appl
c 38	36.4	1.0	600	5	PCT-US96-06352-104	Sequence 104, Appl
c 39	36.4	1.0	1700	5	PCT-US92-02091-1	Sequence 1, Appl
c 40	36.4	1.0	246240	2	US-08-724-394A-20	Sequence 20, Appl
c 41	36.4	1.0	246240	2	US-08-724-394A-21	Sequence 21, Appl
c 42	36.4	1.0	246240	2	US-08-724-394A-22	Sequence 22, Appl
c 43	36.2	1.0	360	4	US-08-118-200-12	Sequence 12, Appl
c 44	36.2	1.0	360	4	US-08-458-745-12	Sequence 12, Appl
c 45	36.2	1.0	2980	2	US-08-463-081B-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match 2.2%; Score 77.4; DB 1; Length 7218;
Best Local Similarity 2.3%; Pred. No. 8.3e-12;
Matches 9; Conservative 245; Mismatches 131; Indels 0; Gaps 0;
QY 2830 gaatttcctattatgagctagcttctggtccatctctggtctctctctgacacatat 2889
Db 1064 GATYVV 1123
QY 2890 tagcttctagccttcttccacgactttattcttcttccacacacatgcttaccacat 2949
Db 1124 YVV 1183
QY 2950 ctctctctgctgtgcttggacttcccccacacagaatttcaacgactctcaagtcttt 3009
Db 1184 YVV 1243
QY 3010 tcttcaccccccac 3069
Db 1244 YVV 1303
QY 3070 gatgctgctatggcctaatattgctttagatgaacattagatttaagcttaagagg 3129
Db 1304 YVV 1363
QY 3130 ttcaaatccac 3189
Db 1364 YVV 1423
QY 3190 actgattgactgaacagagtggtcc 3214
Db 1424 YVV 1488

RESULT 2
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgt-Fls
; US-08-232-463-14
Query Match 1.6%; Score 57.4; DB 1; Length 7218;
Best Local Similarity 4.3%; Pred. No. 4e-06;
Matches 16; Conservative 211; Mismatches 142; Indels 0; Gaps 0;
QY 444 aggaagcaaaaaggaagcacagacagatccctgggagaaaatgccgcgcctctgggt 503
Db 1407 RRR 1348
QY 504 catcgatgagcctcgccctgctgctccctgtgagggaggaagacattagaaaatga 563
Db 1347 RRR 1288
QY 564 attgatgtcttccctaaagatggcagagaaacacagatcctgtgtggatatatttga 623
Db 1287 RRR 1228
QY 624 acgggattacagattgaaatgaagtcacaaagtgcattaccattgagagaaaacag 683
Db 1227 RRR 1168
QY 684 acgagaaaatcttggcttccacagacatgcacaaacacaaatgaatactgtgatga 743
Db 1167 RRR 1108
QY 744 catgaggcagcaagctggggagagagataaccaggggagaggtcagagattctggccc 803
Db 1107 RRR 1048
QY 804 tgctgctca 812
Db 1047 TGCAGCCAA 1039
RESULT 3
US-08-680-327-2/C
; Sequence 2, Application US/08680327
; Patent No. 5859321
; GENERAL INFORMATION:
; APPLICANT: Staskawicz, Brian S., Oldroyd, Giles Edward,
; APPLICANT: Salmemon, John M., Rommens, Caius
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT
; TITLE OF INVENTION: PATHOGEN RESISTANCE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist Sparkman Campbell Leigh &
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,327
; FILING DATE: July 11, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8300)..(9466)
US-09-228-246-1

Query Match          1.2%;   Score 43;   DB 4;   Length 10968;
Best Local Similarity 50.2%;   Pred No, 0.064;
Matches 132; Conservative 0; Mismatches 130; Indels 1; Gaps 1;

QY 1490 ccataccaccctaaacaaaatcttaactgtatcttgaacatgtcaggacatacatat 1549
      |||||      ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 6730 CCATTTCCTTAGGGAAGCTATAACTTGAAGAAATCTGGTAC-TTTTCAGACAAATAATCTT 6672

QY 1550 tcctctgcctcgagaagctcttccctgtctcttaaatctagaatgatgaagtttgaa 1609
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 6671 TACCTCTGCCTCAACAACCTTAAGCAANTGCTTGACATCAGAGAGATATAGTGTGTAATA 6612

QY 1610 taagttgactatcttaacttcattgcagaagaaggagacacatatgagattcatcatcacatga 1669
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6611 CCAAGCAGGAAGTAATAACCCCTTGACAGAGAAACAAATGTAGTTTTATCTTCAACACG 6552

```

Qy 1670 gaagcgaaataactaaaagtgttaatttgattatagaagatttagataaaatatatagaatgca 1799
||||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 6551 TTCAGCGGAGATCTGTGAAGTTCTATATGCATATTCCGGTGACTGTACAATAATCTACAAAGTAA 6492

QY 1730 agagccacagagggaatgtttat 1752
||||||| | | |||||

Db 6491 AGAGCCAATGTGCAATAGTTTTGT 6469

RESULT . 5
US-08-772-440-33/c
; Sequence 33, Application US/08772440
; Patent No. 6046158
; GENERAL INFORMATION:
; APPLICANT: Ariizumi, Kiyoshi
; APPLICANT: Takashima, Akira
; TITLE OF INVENTION: UNIQUE DENDRITIC CELL-ASSOCIATED C-TYPE
; ; TITLE OF INVENTION: LECTINS, DECTIN-1 AND DECTIN-2; COMPOSITIONS AND USES
; ; TITLE OF INVENTION: THEREOF
; ; NUMBER OF SEQUENCES: 42
; ; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433

STATE: TEXAS
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/772,440
FILING DATE: CONCURRENTLY HERewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: OTXD:493
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 10409 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:

Qy 2086 tctgtgtgtgtgtgtgtgtgtgtacatgccaaagtgtgcctctctcttga 2145

[illegible]

```

RESULT 9
US-09-016-366A-13
; Sequence 13, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
; TITLE OF INVENTION: INHIBITORS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,366A
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/037,090
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3757 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-016-366A-13

```

```

Query Match      1.18; Score 40.6; DB 2; Length 3757;
Best Local Similarity 69.6%; Pred. No. 0.17;
Matches 55; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY   2086  tcgtgctgtgtgtgtgtgtgtgcagtcacatgccaaagtgcctctctcttga 2145
Db    1321  TGTGTGTCGTGTGTGTGTGTGTGTGTATGTGTGTGTGTACTGGT 1380

QY   2146  ccattatttcagacttaa 2164
Db    1381  CCCTCATTTGTGCACCAA 1399

RESULT 10
US-08-978-404B-19
; Sequence 19, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
```

[illegible]

```

> PC7-RESU95-04589-157/c
> Sequence 157, Application PC7US9504589
> GENERAL INFORMATION:
> APPLICANT: Ausubel, Frederick M.
> APPLICANT: Staszewicz, Brian J.
> APPLICANT: Brent, Andrew F.
> APPLICANT: Dahlbeck, Douglas
> APPLICANT: Katagiri, Fumitaki
> APPLICANT: Kunkel, Barbara N.
> APPLICANT: Mindrinos, Michael N.
> APPLICANT: Yu, Guo-Liang
> TITLE OF INVENTION: RPS2 GENE AND U
> NUMBER OF SEQUENCES: 201
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: Fish & Richardson
> STREET: 225 Franklin Street Suite
> CITY: Boston
> STATE: MA
> COUNTRY: USA

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:26:25 ; Search time 478.87 Seconds
(without alignments)
4696.773 Million cell updates/sec

Title: US-09-402-713A-6

Perfect score: 3582

Sequence: 1 acagagaataagcaagtgc.....tgattctttgttacaacttt 3582

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq.0601.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3580.4	100.0	3582	19 AAV62430	Prostate cancer an
2	2032	56.7	2037	19 AAV62427	Prostate cancer an
3	1772.4	49.5	1872	19 AAV62428	Prostate cancer an
4	1742.2	48.6	3112	21 AAA06687	Human immunogenic
5	1733.4	48.4	2426	21 AAA06689	Human immunogenic
6	1717.4	47.9	2229	21 AAA06688	Human immunogenic
7	725.4	20.3	820	19 AAV62429	Prostate cancer an
8	722.4	20.2	812	21 AAA06690	Human immunogenic
9	524	14.6	597	20 AAX37486	Human secreted pro
10	457.2	12.8	718	21 AAA06545	Human immunogenic
11	319.6	8.9	437	21 AAC06768	Human secreted pro

c 12	288.4	8.1	301	21	AAA06520	Human immunogenic
c 13	247	6.9	283	21	AAA06468	Human immunogenic
c 14	235.4	6.6	359	20	AAZ33445	Human prostate can
c 15	133	3.7	936	22	AAF58252	Oligonucleotide D1
c 16	133	3.7	936	22	AAF58254	Oligonucleotide D1
c 17	133	3.7	936	22	AAF58257	Oligonucleotide D1
c 18	133	3.7	936	22	AAF58259	Oligonucleotide D2
c 19	133	3.7	936	22	AAF58262	Oligonucleotide D2
c 20	133	3.7	936	22	AAF58255	Oligonucleotide D1
c 21	132.6	3.7	936	22	AAF58255	Oligonucleotide D1
c 22	132.6	3.7	936	22	AAF58254	Oligonucleotide D1
c 23	132.6	3.7	936	22	AAF58257	Oligonucleotide D1
c 24	132.6	3.7	936	22	AAF58259	Oligonucleotide D2
c 25	132.6	3.7	936	22	AAF58262	Oligonucleotide D2
c 26	132.6	3.7	936	22	AAF58255	Oligonucleotide D1
c 27	121	3.4	123	19	AAV33791	Prostate cancer an
c 28	52.2	1.5	244	22	AAF58238	Oligonucleotide D1
c 29	51.4	1.4	244	22	AAF58238	Oligonucleotide D1
c 30	44.2	1.2	3425	15	AAQ57072	AGE-modified DNA I
c 31	43	1.2	10968	19	AAV17789	Tomato Prf genomic
c 32	42.4	1.2	3282	21	AAAS3887	Murine llyby recep
c 33	41.6	1.2	1287	21	AAAS3887	Human secreted pro
c 34	41.6	1.2	10409	19	AAV42558	Mouse dectin-2 gen
c 35	41.4	1.2	5475	19	AAV17777	Tomato Prf cDNA.
c 36	40.8	1.1	319	20	AAV90110	EST clone DG329.
c 37	40.6	1.1	278	18	AAAT65703	Polymorphic repeat
c 38	40.6	1.1	3757	19	AAV44332	Murine mMCP-6 zymo
c 39	40.6	1.1	3757	19	AAV42708	Mouse mast cell pr
c 40	40.6	1.1	43795	21	AAZ92583	Human DAZ genomic
c 41	40.2	1.1	1729	20	AAZ86275	DNA encoding a hum
c 42	40	1.1	796	20	AAZ37484	Human secreted pro
c 43	40	1.1	3710	18	AAV74716	Staphylococcus aur
c 44	39.4	1.1	1660	22	AAF81244	Anthranilate synth
c 45	39.2	1.1	541	21	AAA74269	Loiblolly pine SSR

ALIGNMENTS

RESULT 1
AAV62430
ID AAV62430 standard; cDNA; 3582 BP.
XX
AC AAV62430;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) wild-type cDNA.
XX
KW Prostate cancer antigen cDNA; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 401..556
FT /*tag= a
FT /product= "PCA3 protein"
FT polyA_signal 983..987
FT /*tag= b
FT polyA_signal 2041..2046
FT /*tag= c
FT polyA_signal 2597..2602
FT polyA_signal 3494..3496
FT /*tag= d
FT /*tag= e
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX

Qy 1621 tcttacttcatgcaagaagggacacatatagagatttcattcatcatcatgagacagcaata 1680
Db 1621 tcttacttcatgcaagaagggacacatatagagatttcattcatcatcatgagacagcaata 1680
Qy 1681 ctaaaagtgttaattgatttaagaagttagataaaatatagaatgcaagagccacaga 1740
Db 1681 ctaaaagtgttaattgatttaagaagttagataaaatatagaatgcaagagccacaga 1740
Qy 1741 gggaaatgtttatggggcagctttgaaacctgggatgtgaagcaaaagcagggaacctca 1800
Db 1741 gggaaatgtttatggggcagctttgaaacctgggatgtgaagcaaaagcagggaacctca 1800
Qy 1801 tagtatcttataataatacttcatcttctctctctctctctctctctctctctctctct 1860
Db 1801 tagtatcttataataatacttcatcttctctctctctctctctctctctctctctct 1860
Qy 1861 cacagaattcatgcatgcaaatccccaaaggtaaaccttttccatttcaatggtgagtc 1920
Db 1861 cacagaattcatgcatgcaaatccccaaaggtaaaccttttccatttcaatggtgagtc 1920
Qy 1921 gctttagaattttggcaaatcatactggtcacttctctctctctctctctctctctct 1980
Db 1921 gctttagaattttggcaaatcatactggtcacttctctctctctctctctctctctct 1980
Qy 1981 ttgtagtttaattgaagaataagggcactcttctgtgagccacttttagggttcaactctggc 2040
Db 1981 ttgtagtttaattgaagaataagggcactcttctgtgagccacttttagggttcaactctggc 2040
Qy 2041 aataaagaatttacaaaagactactcaggaccagttgttaagagctctgtgtgtgtgt 2100
Db 2041 aataaagaatttacaaaagactactcaggaccagttgttaagagctctgtgtgtgtgt 2100
Qy 2101 gtgtgtgtgtgagtgatcagccaaagtgtccctctctctctctctctctctctctct 2160
Db 2101 gtgtgtgtgtgagtgatcagccaaagtgtccctctctctctctctctctctctctct 2160
Qy 2161 ttaaaacaagcatgttttcaaatggcactatgagtcgcaatgtatgtatcaccaccat 2220
Db 2161 ttaaaacaagcatgttttcaaatggcactatgagtcgcaatgtatgtatcaccaccat 2220
Qy 2221 ctcatcttctccagtaaatgtgataaataatgtctctctctctctctctctctctctct 2280
Db 2221 ctcatcttctccagtaaatgtgataaataatgtctctctctctctctctctctctctct 2280
Qy 2281 ttcacaaagcagctggaaatggacacacacacacacacacacacacacacacacacac 2340
Db 2281 ttcacaaagcagctggaaatggacacacacacacacacacacacacacacacacacac 2340
Qy 2341 gctcacactgtttgacatatattgttagaagcactctgcattgtgggttctctcttaagc 2400
Db 2341 gctcacactgtttgacatatattgttagaagcactctgcattgtgggttctctcttaagc 2400
Qy 2401 aaaaactctgattagttcagctcagctgggctgtgcatcagcggtttgagaaaattcaa 2460
Db 2401 aaaaactctgattagttcagctcagctgggctgtgcatcagcggtttgagaaaattcaa 2460
Qy 2461 ttctcagcagagcagaaatttgaattccctctctctctctctctctctctctctctct 2520
Db 2461 ttctcagcagagcagaaatttgaattccctctctctctctctctctctctctctctct 2520
Qy 2521 gagaggattcagacagctcagggtgtttcactaatgtctgaaactctgttccctctttg 2580
Db 2521 gagaggattcagacagctcagggtgtttcactaatgtctgaaactctgttccctctttg 2580
Qy 2581 tgttcatgtagtccaataataatgtttatcttttgaactgatgctcatagagagaata 2640
Db 2581 tgttcatgtagtccaataataatgtttatcttttgaactgatgctcatagagagaata 2640
Qy 2641 taagaactctgagtgatatacaacattagggattcaagaataatttagattttaagctcaca 2700
Db 2641 taagaactctgagtgatatacaacattagggattcaagaataatttagattttaagctcaca 2700

Qy 2701 ctgggtcaaaaagaaaccaagatatacaagaactctgagctgtcatcgtcccatctctgtga 2760
Db 2701 ctgggtcaaaaagaaaccaagatatacaagaactctgagctgtcatcgtcccatctctgtga 2760
Qy 2761 gccacaaaccaacagcaggacccaacgcgcatgtctgagatctcttaataaaggaacagtg 2820
Db 2761 gccacaaaccaacagcaggacccaacgcgcatgtctgagatctcttaataaaggaacagtg 2820
Qy 2821 tcatgagttgaattctctctattatgagctgagcttctgagcttctggccatctctgctctctt 2880
Db 2821 tcatgagttgaattctctctattatgagctgagcttctgagcttctggccatctctgctctctt 2880
Qy 2881 gacacatatagcttctgagcttctgcttctccacgacttttattcttttccacacatgcg 2940
Db 2881 gacacatatagcttctgagcttctgcttctccacgacttttattcttttccacacatgcg 2940
Qy 2941 ttaccaaactctctctctctctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3000
Db 2941 ttaccaaactctctctctctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3000
Qy 3001 caagtcttttct 3060
Db 3001 caagtcttttct 3060
Qy 3061 ttccaatagatgctgctctatgggctaattattgcttttagatgaacatttagatttaaag 3120
Db 3061 ttccaatagatgctgctctatgggctaattattgcttttagatgaacatttagatttaaag 3120
Qy 3121 tetaagaggttcaaaaatccaaactctctctctctctctctctctctctctctctctctct 3180
Db 3121 tetaagaggttcaaaaatccaaactctctctctctctctctctctctctctctctctctct 3180
Qy 3181 cctatatattactgattgactgaacagaggtgtccccaagatgccagtcacaaatgagaacc 3240
Db 3181 cctatatattactgattgactgaacagaggtgtccccaagatgccagtcacaaatgagaacc 3240
Qy 3241 cagtggtctctgtggtatcatgcatgcaagactgctgaagccagaggtgactgattacg 3300
Db 3241 cagtggtctctgtggtatcatgcatgcaagactgctgaagccagaggtgactgattacg 3300
Qy 3301 cctcatgtgggtggagggagccactcctggcctctgattgttcaggagcagacctgaga 3360
Db 3301 cctcatgtgggtggagggagccactcctggcctctgattgttcaggagcagacctgaga 3360
Qy 3361 tgcctctctctcagtgct 3420
Db 3361 tgcctctctctcagtgct 3420
Qy 3421 ctacattttgagaattcccaattaggaaactcacaatgttttatctgcccctatcaatttttaa 3480
Db 3421 ctacattttgagaattcccaattaggaaactcacaatgttttatctgcccctatcaatttttaa 3480
Qy 3481 acttctgaaaaaataagtttttttcaaaaatctgctctgttaaatattacttttttctacagtg 3540
Db 3481 acttctgaaaaaataagtttttttcaaaaatctgctctgttaaatattacttttttctacagtg 3540
Qy 3541 tcttgagcactatatacaactttgattttgtttacaacttt 3582
Db 3541 tcttgagcactatatacaactttgattttgtttacaacttt 3582

RESULT 2
AAV62427 standard; cDNA; 2037 BP.
XX
AC AAV62427;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 1.
XX
KW Prostate cancer antigen cDNA splice variant 1; PCA3; prostatic cancer;
PC; ds.

QY 821 gttcataaccaaatcttttcatatttttaacccctcaaaacaaagctgtgttaatatctga 880
Db |||||||
QY 634 gttcataaccaaatcttttcatatttttaacccctcaaaacaaagctgtgttaatatctga 693
Db |||||||
QY 881 tctctacggttctcttctggcccaaacattctcaatatataccagccacactcattttta 940
Db |||||||
QY 694 tctctacggttctcttctggcccaaacattctcaatatataccagccacactcattttta 753
Db |||||||
QY 941 atttagtctccagatctgactgtgacctttctacactgtagataaacattactcatttt 1000
Db |||||||
QY 754 atttagtctccagatctgactgtgacctttctacactgtagataaacattactcatttt 813
Db |||||||
QY 1001 gttcaagaacccctctgttctgctcctaataatgtactgttttttcttaaggagtgt 1060
Db |||||||
QY 814 gttcaagaacccctctgttctgctcctaataatgtactgttttttcttaaggagtgt 873
Db |||||||
QY 1061 tctggcccaaggagatctgtgaacagctgggaagcattctcaagatcttccagggttata 1120
Db |||||||
QY 874 tctggcccaaggagatctgtgaacagctgggaagcattctcaagatcttccagggttata 933
Db |||||||
QY 1121 ctctactgacacagcatgatcattacgagtgaaattatctcaatcaacatcatcctcagt 1180
Db |||||||
QY 934 ctctactgacacagcatgatcattacgagtgaaattatctcaatcaacatcatcctcagt 993
Db |||||||
QY 1181 gtcttggccataactgaaattctcccaacttttggccattctcaagacctcaaaa 1240
Db |||||||
QY 994 gtcttggccataactgaaattctcccaacttttggccattctcaagacctcaaaa 1053
Db |||||||
QY 1241 gtcattccattataacacaggaattaaactttttttttaaacctgggaagattcaatgtta 1300
Db |||||||
QY 1054 gtcattccattataacacaggaattaaactttttttttaaacctgggaagattcaatgtta 1113
Db |||||||
QY 1301 catcgagctatgggaatttaattacataattttgttctcagtgcaagatgactaagctc 1360
Db |||||||
QY 1114 catcgagctatgggaatttaattacataattttgttctcagtgcaagatgactaagctc 1173
Db |||||||
QY 1361 ttatccctccctctgttggatttttttccagtaataaagttaaactgttagccttgt 1420
Db |||||||
QY 1174 ttatccctccctctgttggatttttttccagtaataaagttaaactgttagccttgt 1233
Db |||||||
QY 1421 actgagctgtatcacagcagcctccctccctccctccctccctccctccctccctccat 1480
Db |||||||
QY 1234 actgagctgtatcacagcagcctccctccctccctccctccctccctccctccctccat 1293
Db |||||||
QY 1481 caacctctccatcacacataacaaaacttaactgttaattctctgaacatgtcaggac 1540
Db |||||||
QY 1294 caacctctccatnyaaactaaacaaaacttaactgttaattctctgaacatgtcaggnc 1353
Db |||||||
QY 1541 atacattatctctgctggaagctctctctctctctctctctctctctctctctctctct 1600
Db |||||||
QY 1354 atacatttctctgctggaagctctctctctctctctctctctctctctctctctctct 1413
Db |||||||
QY 1601 agtttgaataagtgaacttacttacttacttacttacttacttacttacttacttacttact 1660
Db |||||||
QY 1414 agtttgaataagtgaacttacttacttacttacttacttacttacttacttacttacttact 1473
Db |||||||
QY 1661 atcacatgagacgcaaaactactaaaagtgttaatttgaataagagtttagataaata 1720
Db |||||||
QY 1474 atcacatgagacgcaaaactactaaaagtgttaatttgaataagagtttagataaata 1533
Db |||||||
QY 1721 tgaattgcaagagccacagaggaaagtgttatggggcagctgtttgtaagcctggagtga 1780
Db |||||||
QY 1534 tgaattgcaagagccacagaggaaagtgttatggggcagctgtttgtaagcctggagtga 1593
Db |||||||
QY 1781 agcaaaagcagggaacccatagatcttataataataacttacttctctctctctctctct 1840
Db |||||||
QY 1594 agmaaaagcagggaacccatagatcttataataataacttacttctctctctctctctct 1653
Db |||||||
QY 1841 acaatatcaaacagcttttccagaaattctagcagtgcaaatcccaaggttaaccttt 1900
Db |||||||
QY 1654 acaatatcaaacagcttttccagaaattctagcagtgcaaatcccaaggttaaccttt 1713
Db |||||||

QY 1901 atocatttcagtgagtgagtcgcttttagaatttttggcaaatcactactgttcacttatctca 1960
Db |||||||
QY 1714 atocatttcagtgagtgagtcgcttttagaatttttggcaaatcactactgttcacttatctca 1773
Db |||||||
QY 1961 actttgagatggtttgtctctgttagttaattgaaagaataaggcactctgttgagcca 2020
Db |||||||
QY 1774 actttgagatggtttgtctctgttagttaattgaaagaataaggcactctgttgagcca 1833
Db |||||||
QY 2021 ctttagggttctactcctcctggaataaaagaattttacaaga 2059
Db |||||||
QY 1834 ctttagggttctactcctcctggaataaaagaattttacaaga 1872
Db |||||||

RESULT 4

AAA06687
ID AAA06687 standard; cDNA; 3112 BP.

XX AC AAA06687;

XX XX 13-JUN-2000 (first entry)

DT DT

XX XX Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:468.

DE Human; prostate cancer; diagnosis; tumour; gene therapy; detection;

XX Human immunogenic; cytostatic; vaccine; ss.

KW Homo sapiens.

XX OS

XX WO200004149-A2.

PN XX

XX PD 27-JAN-2000.

XX PF 14-JUL-1999; 99WO-US15838.

XX PR 14-JUL-1998; 98US-0115453.

XX PR 14-JUL-1998; 98US-0115453.

XX PR 23-SEP-1998; 98US-0159812.

XX PR 23-SEP-1998; 98US-0159812.

XX PR 15-JAN-1999; 99US-0232149.

XX PR 15-JAN-1999; 99US-0232880.

XX PR 09-APR-1999; 99US-0288946.

XX PA (CORI-) CORIXA CORP.

XX PI Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;

XX DR WPI; 2000-171268/15.

XX PT New polypeptide useful for treating and diagnosing prostate cancer

XX PS comprises an immunogenic portion of prostate tumor protein .

XX CC Claim 1; Page 259-260; 263pp; English.

XX CC The present invention describes isolated polypeptides, comprising an

XX CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides

XX CC and polynucleotides encoding them have cytostatic activity and can be

XX CC used in vaccines and in gene therapy. The polypeptides and

XX CC polynucleotides encoding them, antigen presenting cells which express

XX CC the polypeptides, antibodies against the polypeptides and vaccines

XX CC comprising them can be used for inhibiting the development of prostate

XX CC cancer in a patient. The polypeptides can be used to generate antibodies

XX CC or anti-idiotypic antibodies for passive immuno therapy. A portion of

XX CC the polynucleotides encoding the polypeptides can be used as a probe or

XX CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and

XX CC AAY82000 to AAY82020 represent sequences used in the exemplification of

XX CC the present invention.

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

Query Match 48.6%; Score 1742.2; DB 21; Length 3112;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1780; Conservative 0; Mismatches 8; Indels 3; Gaps 3;

Qy	274	tcaattggccagggtgagaaataagaaaggctgctgaactttaccatctcgagggccacacatc	333
Db	1302	tsactcaaatagggtgagaaataagaaaggctgctgaactttaccatctcgagggccacacatc	1361
Qy	334	tgcagaattggagataaattaaacatcacctagaaacagcaagatgacaataaaatgtctaaag	393
Db	1362	tgtgaaatggagataaattaaacatcacctagaaacagcaagatgacaataaaatgtctaaag	1421
Qy	394	tagtgacatgttttgcacatttcagccctttaaataatccacacacagagaacaca	453
Db	1422	tagtgacatgttttgcacatttcagccctttaaataatccacacacagagaacaca	1481
Qy	454	aaaggagcacagagatccctggggaaatcccgccgcgcacatcttggttcacatcgatgag	513
Db	1482	aaaggagcacagagatccctggggaaatcccgccgcgcacatcttggttcacatcgatgag	1541
Qy	514	cctcgccctgtgccctgtcccgcttgtgaggaagacattagaaaaatgaattgattgtgt	573
Db	1542	cctcgccctgtgccctgtcccgcttgtgaggaagacattagaaaaatgaattgattgtgt	1601
Qy	574	tcctaaaggatggcaggaanaacagatccctgttgtgatatatttatgtgaacgggattac	633
Db	1602	tcctaaaggatggcaggaanaacagatccctgttgtgatatatttatgtgaacgggattac	1661
Qy	634	agatttgaattgaagtccaagaatgagcattaccaatgagaggaaaaacagacagaaaaat	693
Db	1662	agatttgaattgaagtccaagaatgagcattaccaatgagaggaaaaacagacagaaaaat	1721
Qy	694	cttgatggcttcacaagacatgcaacaacaaaaatggaaatctgtgatgacatgaggcag	753
Db	1722	cttgatggcttcacaagacatgcaacaacaaaaatggaaatctgtgatgacatgaggcag	1781
Qy	754	ccaagctgggaggagataaacacagggcagagggctcagattctgcccctgtgcctaa	813
Db	1782	ccaagctgggaggagataaacacagggcagagggctcagattctgcccctgtgcctaa	1841
Qy	814	actgtgcgttcataacaaatcatttcatttctaocctcacaacaaagcgttgtgttaa	873
Db	1842	actgtgcgttcataacaaatcatttcatttctaocctcacaacaaagcgttgtgttaa	1901
Qy	874	tatctgatctctacggttccctctgtgggcccacaattctccaatatccagccacacatcat	933
Db	1902	tatctgatctctacggttccctctgtgggcccacaattctccaatatccagccacacatcat	1961
Qy	934	ttttaatatattagttccagatctgactgtgaccttctacactgtagataaacattac	993
Db	1962	ttttaatatattagttccagatctgactgtgaccttctacactgtagataaacattac	2021
Qy	994	tcattttgttcaaaagacctctgtgtgctgcctaaatagttagctgactgttttccctaa	1053
Db	2022	tcattttgttcaaaagacctctgtgtgctgcctaaatagttagctgactgttttccctaa	2081
Qy	1054	ggagtgcttctggcccaggggatactgtgaacagctgggaaacatctcaagatctttccag	1113
Db	2082	ggagtgcttctggcccaggggatactgtgaacagctgggaaacatctcaagatctttccag	2141
Qy	1114	ggttatacttactagcaacacagatgatacttaaggagtgaaattatctaatcaacatcat	1173
Db	2142	ggttatacttactagcaacacagatgatacttaaggagtgaaattatctaatcaacatcat	2201
Qy	1174	cctcagtgcttttgcccaactgagaaattcatttcccacttttgtgcccattctcaagacc	1233
Db	2202	cctcagtgcttttgcccaactgagaaattcatttcccacttttgtgcccattctcaagacc	2261
Qy	1234	tcaaaatgtcattcccaataatcacaggattaaacttttttttaacctgggaagaattc	1293
Db	2262	tcaaaatgtcattcccaataatcacaggattaaacttttttttaacctgggaagaattc	2321
Qy	1294	aatgttacatcgagctatgggaatttaattacataattttgtttccagtgcaaaagatgac	1353
Db	2322	aatgttacatcgagctatgggaatttaattacataattttgtttccagtgcaaaagatgac	2381

[illegible]

RESULT	5	
AAA06689/c		
ID	AAA06689	standard; cdNA; 2426 BP.
XX	AC	
XX	AC	
XX	AAA06689;	
DT	13-JUN-2000	(first entry)
XX		
DE	Human immunogenic prostate tumour protein	cdNA sequence SEQ ID NO:470.
XX		
XX	Human; prostate cancer; diagnosis; tumour; gene therapy; detection;	
KW	immunogenic; cytostatic; vaccine; ss.	
KW		
XX		
OS	Homo sapiens.	
XX		
PN	WO200004149-A2.	
XX		
PD	27-JAN-2000.	
XX		
PF	14-JUL-1999;	99WO-US15838.
XX		
PR	14-JUL-1998;	98US-0115453.
PR	14-JUL-1998;	98US-0116134.
PR	23-SEP-1998;	98US-0159812.

|||||
Db 151 CATTTTCATGGTGGTGGCGCTTTAGAAATTTGGCAATCATACTGGTCACTTATCTCAACT 92
QY 1964 ttgaagatgtttgtcccttgtagttaattgaagaataggcactcttgtgagcactt 2023
Db 91 TTGAGATGTTGTGCTTGTAGTTAAATTTGAAGAAATAGGGCACTCTTGTGAGCCACTT 32
QY 2024 tagggctcactctggcaataagaatttac 2054
Db 31 TAGGGTTCACTCTCGCAATAAAGAAATTTAC 1
RESULT 6
ID AAA06688/c
XX AAA06688 standard; CDNA; 2229 BP.
AC AAA06688;
DT 13-JUN-2000 (first entry)
XX Human immunogenic prostate tumour protein CDNA sequence SEQ ID NO:469.
DE Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX Homo sapiens.
XX WO200004149-A2.
PN 27-JAN-2000.
PD 14-JUL-1999; 99WO-US15838.
PF 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0115453.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX (CORI-) CORIXA CORP.
XX Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
PI WPI; 2000-171268/15.
XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
PS Claim 1; Page 260-261; 263pp; English.
XX The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides; AAA06241 to AAA06691 and
CC AAA82000 to AAA82020 represent sequences used in the exemplification of
XX the present invention.
SQ Sequence 2229 BP; 654 A; 447 C; 481 G; 647 T; 0 other;

Query Match 47.9%; Score 1717.4; DB 21; Length 2229;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1773; Conservative 0; Mismatches 1; Indels 5; Gaps 5;

QY 285 ggtgagaataaagaagctgctgactttaccatctgagggcacacatctgctgaaatgg 344
Db 1775 GGTGAGAAATAAGAAAGGCTGCTGACTTTACCATCTGAGGCCACACATCTGCTGAATGG 1716
QY 345 agataataacatcactactagaacacagacagatgaacaataatgtctaaagttagcatgt 404
Db 1715 AGATAATTAACATCACTAGAAAACAGCAAGATGACAATAATTAATGTCTTAAGTAGCATGT 1656
QY 405 ttttgacatttccagccctttaaataatccacacacacagagaagcaaaaggaagcac 464
Db 1655 TTTTGCACATTTCCAGCCCTTTAAATATCCACACACAGGAGACACAAAAGGAGACAC 1596
QY 465 agagatccctgggagaataatgcccgcgcctcttgggtcatcgatgagcctcgcctgt 524
Db 1595 AGAGATCCCTGGGAGAAATGCCGCCGCATCTTTGGGTATCGATGAGCCCTCGCCCTGT 1536
QY 525 gctgtgtccctgtgtgaggggaagacattagaaaatgaatgatgtgttccctaaagga 584
Db 1535 GCCTGTCCCTGTGTGAGGGAAGGACATTTAGAAAATGAATTTGATGTCTTCTTTAAGGA 1476
QY 585 tggcagagaaaacagatcctgtgttggtatttttgaacgggattacagatttgaat 644
Db 1475 TGGGAGGAAAACAGATCCTGTGTGGATATTTATTGAACGGGATACAGATTTGAAT 1416
QY 645 gaagtcaaaaagtgcagcattaccatgcagagaaaacagacagaaaaacttgcgtgctt 704
Db 1415 GAAGTCACAACTGAGCATTTACCAATGAGGGAACACACAGCAAAATCTTTGATGGCTT 1356
QY 705 cacaagacatgcacaaacaaaatggaatactgtgatgcacatgagggcagcgaagtgggg 764
Db 1355 CACAAGACATGCAACAAACAAAATGGAATCTGTGATGACATGAGCGACGCCAAGCTGGG 1296
QY 765 agagataaaccacggggcagaggttcaggattctggccctgctgctaaactgtcgcttc 824
Db 1295 AGGAGATAACCCACGGGGCAGAGGGTCAGGATTTCTGGCCCTGTGCTAAACCTGGGTTTC 1236
QY 825 ataaccaaatcttcatatttctaacctcctcaaaacaaagctgtgtgtaatactcgtatcc 884
Db 1235 ATAACCAATCATTTTCATATTTCTAACCTCAAAACAAAGCTGTGTGTAATATCTGATCTC 1176
QY 885 taccgttcccttctggggcccaacattctccatatatccagccacactcatcttataattt 944
Db 1175 TAGGGTTCTTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCATTTTAAATATT 1116
QY 945 agttccagatctgtactgtaccttctacacttagaataaacattactctatttcttc 1004
Db 1115 AGTCCAGATCTGACTGTGACCTTTTACACTGTAGAAATACATTAATCTCAATTTGTTTC 1056
QY 1005 aaagaccctctgtgtgtgcctaataatagctgactgttttccctaaggaggtgtctg 1064
Db 1055 AAAGACCCTTCTGTGTGCTGCCTAATATATCTAGCTGACTGTCTTTTCTTAAGGAGTGTCTG 996
QY 1065 gccaggggagctgtgaacaggtgggaagcatctcaagatcttccaggggtatactta 1124
Db 995 GCCCAGGGGATCTGTGACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGGTTATACTTGA 936
QY 1125 ctgacacacagatcaatcattacagagtgaaattatctaataacatccactcagtgct 1184
Db 935 CTAGCACACAGGATGATCATTTACGGAGTGAAATATCTAATCAACATCATCTCCAGTGTCT 876
QY 1185 ttgccataactgaattctccacttttgcaccttctcaagacctcaaaatgtca 1244
Db 875 TTGCCATFACTGAATTCATTTCCCACTTTTGTGCCCACTTCTCAAGACCTCAAAATGTCA 816
QY 1245 ttccattataatacacaggattaaacttttttttaacctgggaagaaattcaatttcatatg 1304
Db 815 TTCCATTAATATCACAGGATTAACTTTTAACTTGAAGAAATTCATGTTTACATG 757
QY 1305 cagcatatgggaatttaattacatatatttcttccagtgcaaaagatgactaaagtccttta 1364
Db 756 CAGCATTTGGGAATTTAAATTTACATATTTTGTCTTCCAGTGCAAAAGATGACTAAAGTCCTTTA 697
QY 1365 tccctcccttctgttgatttttttttccagtgataaaagttaaaatgcttagccttgactg 1424

QY 821 gttcataaacaatcatttcatttctaaacctcaaaacagctgtgtgtaatactga 880
Db 634 gttcataaacaatcatttcatttctaaacctcaaaacagctgtgtgtaatactga 693
QY 881 tcttacagggttctctctgggcccacattctccatatatccagccacactcaattttta 940
Db 694 tcttacagggttctctctgggcccacattctccatatatccagccacactcaattttta 753
QY 941 atttagttccagatctgactgtgaccttctacactgtagaataaacattactcatttt 1000
Db 754 atttagttccagatctgactgtgaccttctacactgtagaataaacattactcatttt 813
QY 1001 gttcaaa 1007
Db 814 gttcaaa 820

RESULT 8
ID AAA06690/c
XX AAA06690 standard; cDNA; 812 BP.
AC AAA06690;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:471.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 98US-0232149.
PR 15-JAN-1999; 98US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS
PS Claim 1; page 262; 263pp; English.

XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAA82000 to AAA82020 represent sequences used in the exemplification of
CC the present invention.

XX
SQ Sequence 812 BP; 198 A; 189 C; 169 G; 256 T; 0 other;

Query Match 20.2%; Score 722.4; DB 21; Length 812;
Best Local Similarity 99.9%; Pred. No. 2.8e-188;
Matches 723; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 281 caggggtgagaaataagaagagctgactttacacatctgagccacacatctgctgaa 340
Db 724 CAGAGGTGAGAAATAGAAAGGCTGCTGACTTTTACCATCTGAGGCCACACATCTCTGAA 665
QY 341 atggagataaataacatcactagaacaagacattgacaataataatgctgaagtgaac 400
Db 664 ATGGAGATAAATTAACATCACTAGAACAGCAAGATGACAATAATATGCTTAAGTAGTGAC 605
QY 401 atgttttgcacattccagccctttaaataatccacacacaggaagacacaaaggaa 460
Db 604 ATGTTTTTGCACATTTCCAGCCCCCTTTAAATATCCACACACACAGGAAGCAAAAAGGAA 545
QY 461 gcacagagatccctgggagaaatgcccgccgcctcttgggtcatcgatgagcctcgcc 520
Db 544 GCACAGAGATCCCTGGGAGAAATGCCCGCCGCCTCTTGGGTCTCATGATGAGCCTCGCC 485
QY 521 ctgtgcctgggtcccgcttggaggaagacattgaaataatgaattgattgttctctaa 580
Db 484 CTGTGCTGGTCCCGCTTGTGAGGGAAGGACATTAGAAATGAATGATGTGTCTCTTAA 425
QY 581 aggtgggcaggaacacagatcctgtgtggatatttattgaacgggattacagatttg 640
Db 424 AGGATGGCGCAGGAAACAGATCCTGTGTGGATATTATTTCACCGGGATTACAGATTGG 365
QY 641 aaatgaagtacaaagtgcattaccatgagaggaacacagagagaaaaatcttgatg 700
Db 364 AAATGAAGTCACAAAGTGAGCATTTACCAATGAGAGGAAAAACAGAGAGAAAAATCTTGATG 305
QY 701 gttcacagacatgcaacaaacaaatgaataactgtgactgtgacatgagcgcaagct 760
Db 304 GCTTCACAGACATGCACAAACAAATGGAATCTGTGATGACATGAGGCGAGCCAAAGCT 245
QY 761 ggggaggagataaccacgggagaggggtcaggattcttgcctctgctcctaaactgtgc 820
Db 244 GGGGAGGAGATAACCCAGGGGAGAGGTCAGGATCTTGCCCTGCTGCTAACTGTGC 185
QY 821 gttcataaacaatcatttcatttctaaacctcaaaacagctgtgtgtaatactga 880
Db 184 GTTCATAACCAAAATCATTTTCATATTCTAACCTCAAAACAAAGCTGTTGTAATCTGA 125
QY 881 tcttacagggttctctctgggcccacattctccatatatccagccacactcaatttt 940
Db 124 TCTCTACGGTTCCTTCTGGGCCCAACATTTCTCATATATCCAGCCACACTCAATTTTAAAT 65
QY 941 atttagttccagatctgactgtgaccttctacactgtagaataaacattactcatttt 1000
Db 64 ATTTAGTTCCAGATCTGTACTGTGACCTTTCTACACTGTAGAAATACATTACTCATTTT 5
QY 1001 gttc 1004
Db 4 GTTC 1

RESULT 9
AA37486
ID AA37486 standard; cDNA; 597 BP.
XX
AC AA37486;
XX
DT 06-JUL-1999 (first entry)
XX
DE Human secreted protein cDNA fragment containing gene 36.
XX
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;

KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
XX arthritis; malignancy; digestive; endocrine; infection; ss.

OS Homo sapiens.

PN W09918208-AL.

XX 15-APR-1999.

XX 01-OCT-1998; 98WO-US20775.

XX 02-OCT-1997; 97US-0060884.

XX 02-OCT-1997; 97US-0060833.

PR 02-OCT-1997; 97US-0060836.

PR 02-OCT-1997; 97US-0060837.

PR 02-OCT-1997; 97US-0060838.

PR 02-OCT-1997; 97US-0060839.

PR 02-OCT-1997; 97US-0060843.

PR 02-OCT-1997; 97US-0060862.

PR 02-OCT-1997; 97US-0060866.

PR 02-OCT-1997; 97US-0060874.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;

PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;

PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;

XX WPI: 1999-264022/22.

DR P-PSDB; AAY07887.

XX New isolated human genes and the secreted polypeptides they encode

XX Claim 1a; Page 247; 368pp; English.

XX This invention describes novel isolated human genes and the secreted

CC proteins they encode. The products of the invention are useful for

CC preventing, treating or ameliorating medical conditions, e.g. by protein

CC or gene therapy. Also pathological conditions can be diagnosed by

CC determining the amount of the new polypeptides in a sample or by

CC determining the presence of mutations in the new polynucleotides.

CC Specific uses are described for each of the 101 polynucleotides, based

CC on which tissues they are most highly expressed in, and include

CC developing products for the diagnosis or treatment of cancer, tumours,

CC neurodegenerative disorders, developmental abnormalities and fetal

CC deficiencies, blood disorders, leukemias, diseases of the immune system,

CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,

CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate

CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,

CC transplant rejection, disorders involving osteoclasts such as

CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,

CC infections and AIDS. The human secreted proteins of the invention are

CC represented in AAY07852-Y07993 and the encoding nucleic acids are

XX represented in AAX37451-X37552.

XX Sequence 597 BP; 181 A; 131 C; 134 G; 150 T; 1 other;

Query Match 14.6%; Score 524; DB 20; Length 597;
Best Local Similarity 99.6%; Pred. No. 7.3e-134;
Matches 524; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 484 gcccgccgcacattgggttcacgtatgagcctcgccctgctgctggtcccgctgtgag 543

|||||

Db 54 gcccgccgcacattgggttcacgtatgagcctcgccctgctgctggtcccgctgtgag 113

|||||

QY 544 ggaagacattagaaatgaattgattgttctctaaagatggcgagaaacagatcc 603

|||||

Db 114 ggaagacattagaaatgaattgattgttctctaaagatggcgagaaacagatcc 173

|||||

QY 604 tgttggtgattatttgaacgggattacagatttgaaatgaatgcacaaagtggatc 663

Db 174 tgttggtgattatttgaacgggwtacagatttgaaatgaatgcacaaagtggatc 233

QY 664 taccatgagagaaacacagacagagaaaatcttgatggtctcacaagacatgcacaaac 723

|||||

Db 234 taccatgagagaaacacagacagagaaaatcttgatggtctcacaagacatgcacaaac 293

QY 724 aaaaatgaatactgtgatgacatgagcgagcgaagctggggagagataaccacggggca 783

|||||

Db 294 aaaaatgaatactgtgatgacatgagcgagcgaagctggggagagataaccacggggca 353

QY 784 gagggtcaggattctgcccctgctgctaaactgtcgcttcaataaccacaaatcttcata 843

|||||

Db 354 gagggtcaggattctgcccctgctgctaaactgtcgcttcaataaccacaaatcttcata 413

QY 844 ttctaacctcaaaaacaaagctgttgtaatactgtatctctacgttctctctggccc 903

|||||

Db 414 ttctaacctcaaaaacaaagctgttgtaatactgtatctctacgttctctctggccc 473

QY 904 aacattctccatatatccagccacactcatttttaataatttagttccagatctgtactg 963

|||||

Db 474 aacattctccatatatccagccacactcatttttaataatttagttccagatctgtactg 533

QY 964 tgacctttctacactgtagaataaacattactcattttgtttcaaga 1009

|||||

Db 534 tgacctttctacactgtagaataaacattactcattttgtttcaaaaa 579

RESULT 10

AAA06545

ID AAA06545 standard; cDNA; 718 BP.

XX AAA06545;

XX 13-JUN-2000 (first entry)

DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:313.

KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;

KW immunogenic; cytostatic; vaccine; ss.

XX Homo sapiens.

XX WO200004149-A2.

PN 27-JAN-2000.

XX 14-JUL-1999; 99WO-US15838.

XX 14-JUL-1998; 98US-0115453.

PR 14-JUL-1998; 98US-0116134.

PR 23-SEP-1998; 98US-0159812.

PR 23-SEP-1998; 98US-0159822.

PR 15-JAN-1999; 99US-0232149.

PR 15-JAN-1999; 99US-0232880.

PR 09-APR-1999; 99US-0288946.

XX (CORI-) CORIXA CORP.

XX Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;

XX WPI: 2000-171268/15.

PT New polypeptide useful for treating and diagnosing prostate cancer

XX comprises an immunogenic portion of prostate tumor protein -

PS Claim 1; Page 199-200; 263pp; English.

XX The present invention describes isolated polypeptides, comprising an

CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides

CC and polynucleotides encoding them have cytostatic activity and can be

CC used in vaccines and in gene therapy. The polypeptides and

CC polynucleotides encoding them, antigen presenting cells which express

CC the polypeptides, antibodies against the polypeptides and vaccines

comprising them can be used for inhibiting the development of prostate cancer in a patient. The polypeptides can be used to generate antibodies or anti-idiotypic antibodies for passive immuno therapy. A portion of the polynucleotides encoding the polypeptides can be used as a probe or to modulate the expression of the polypeptides. AA06241 to AA06691 and AA82000 to AY82020 represent sequences used in the exemplification of the present invention.

Sequence 718 BP; 222 A; 145 C; 169 G; 172 T; 10 other;

Query Match 12.8%; Score 457.2; DB 21; Length 718;
Best Local Similarity 97.2%; Pred. NO. 1.8e-115;
Matches 518; Conservative 0; Mismatches 9; Indels 6; Gaps 5;

QY 281 caggggtagaataagaaggctgctgactttaccattgagggccacacatctgtgaa 340
DB |||||
DB 69 cagaggtagaataagaaggctgctgactttaccattgagggccacacatctgtgaa 128
QY 341 atggagataattacatcactagaaacagagatgacaataataatgtctaagttagtgac 400
DB |||||
DB 129 atggagataattacatcactagaaacagagatgacaataataatgtctaagttagtgac 188
QY 401 atgtttttgacatttcacagcccttttaataatccacacacagagacacaaaggaa 460
DB |||||
DB 189 atgtttttgacatttcacagcccttttaataatccacacacagagacacaaaggaa 248
QY 461 gcacagagatccctgggagaaatcccgccacatcttggctcatgagcctgcc 520
DB |||||
DB 249 gcacagagatccctgggagaaatcccgccacatcttggctcatgagcctgcc 308
QY 521 ctgtgctgtgctccctgtgtgagggaagacattagaaatgaattgattgttctctaa 580
DB |||||
DB 309 ctgtgctgtgctccctgtgtgagggaagacattagaaatgaattgattgttctctaa 368
QY 581 aggatgggcaggaacacagatcctgtgtggtattttattttgaacgggattacagatttg 640
DB |||||
DB 369 aggat-ggcaggaacacagatcctgtgtggtattttattttgaacgggattacagatttg 427
QY 641 aaatgaagtcaaaagtgcattaccatgagagaaacagacagagaaatcttgatg 700
DB |||||
DB 428 aaatgaagtcaaaagtgcattaccatgagagaaacagacagagaaatcttgatg 487
QY 701 gcttcacagacatgcaacacaaatggaatactgtgatgacatgagggcagcaagct 760
DB |||||
DB 488 g-ttcacagacatgcaacacaaatggaatactgtgatgacacagag--cagccaact 544
QY 761 gggggaggagataaccacggggcagagggtcaggattctgtgcccctgtgctctaa 813
DB |||||
DB 545 gggggaggagat-accacggggcaga-ggtcaggattctgtgcccctgtgctctaa 595

RESULT 11
AAC06768
ID AAC06768 standard; cDNA; 437 BP.

XX AAC06768;

XX 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 10843.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

XX EP1033401-A2.

PN 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

PF

PR 26-FEB-1999; 99US-0122487.
XX (GEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
DR

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 10843; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.

XX Sequence 437 BP; 140 A; 87 C; 118 G; 88 T; 4 other;

Query Match 8.9%; Score 319.6; DB 21; Length 437;
Best Local Similarity 98.5%; Pred. NO. 8.4e-78;
Matches 319; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 281 cagggtgagaaataagaaggctgctgactttaccattgagggccacacatctgtgaa 340
DB |||||
DB 114 cagaggtagaataagaaggctgctgactttaccattgagggccacacatctgtgaa 173
QY 341 atggagataattacatcactagaaacagagatgacaataataatgtctaagttagtgac 400
DB |||||
DB 174 atggagataattacatcactagaaacagagatgacaataataatgtctaagttagtgac 233
QY 401 atgtttttgacatttcacagcccttttaataatccacacacagagacacaaaggaa 460
DB |||||
DB 234 atgtttttgacatttcacagcccttttaataatccacacacagagacacaaaggaa 293
QY 461 gcacagagatccctgggagaaatcccgccacatcttgggtcatgagcctgcc 520
DB |||||
DB 294 gcacagagatccctgggagaaatcccgccacatcttgggtcatgagcctgcc 353
QY 521 ctgtgctgtgctccctgtgtgagggaagacattgaaatgaattgattgttctctaa 580
DB |||||
DB 354 ctgtgctgtgctccctgtgtgagggaagacattgaaatgaattgattgttctctaa 413
QY 581 aggatgggcaggaacacagatcct 604
DB |||||
DB 414 aggatgggcaggaacacagatcct 437

RESULT 12
AAA06520/c
ID AAA06520 standard; cDNA; 301 BP.

XX AAA06520;

XX 13-JUN-2000 (first entry)

DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:287.

XX Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.

```
XX Homo sapiens.
OS
PN WO200004149-A2.
XX
PD
PF 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 192; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PMP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 301 BP; 76 A; 58 C; 70 G; 97 T; 0 other;

Query Match 8.1%; Score 288.4; DB 21; Length 301;
Best Local Similarity 99.3%; Pred. No. 2.5e-69;
Matches 300; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 659 agcattaccaatgagaggaacacagagagaaatcttgatggcttcacaagacatgcaa 718
Db 301 AGCATTACCAATGAGAGGAAACACAGACGAGAAAAATCTTGATGCTTCACAAAGACATGCAA 242
QY 719 caaacaaaatggaatactgtgatgacatgagcagcagcaagctgggagagataaacag 778
Db 241 CAACAAAATGGAATCTGTGATAACATGAGGACGCAAGCTGGGAGGAGATAAACACG 182
QY 779 ggcagagggctcaggatctgcctgcctgcctaaactgtgccttcataaccacaaatcatt 838
Db 181 GGCAGAGGGTCAAGGATCTGGCCCTGCTGCCCTAAACTGTGCGCTTCATAACCAATCAAT 122
QY 839 tcattattctaacccctcaaaacaaaagctgtgtgaatatctgatctcaggttcctctg 898
Db 121 TCATATTCTTAACCCCTCAAAACAAAGCTGTGTGAATATCTGATCTCTAC-GTTCCCTCTG 63
QY 899 ggcacaaattctccatatatccagccacaaactcatttttaatttagttccagatctg 958
Db 62 GGCCCAACATCTCCATATATCCAGCCACACTCATTATTTAATATTAGTTCCAGATCTG 3
QY 959 ta 960
Db 2 TA 1
```

```
RESULT 13
AAA06468
ID AAA06468 standard; cDNA; 283 BP.
XX
AC AAA06468;
XX
DT 13-JUN-2000 (first entry)
XX
XX Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:235.
DE
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 179; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PMP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 283 BP; 84 A; 50 C; 60 G; 89 T; 0 other;

Query Match 6.9%; Score 247; DB 21; Length 283;
Best Local Similarity 98.6%; Pred. No. 5.7e-58;
Matches 281; Conservative 0; Mismatches 0; Indels 4; Gaps 3;

QY 2425 tggggcctgtgcatcaggc-ggtttgagaaatattcaattctcagagaccagaatttg 2483
Db 1 tggggcctgtgcatcaggcgggtttgagaaatattcaattctcagagaccagaatttg 60
QY 2484 aattccctcatctttka-ggaatcattaccaggtttggagaggattccagagcgtcag 2542
Db 61 aattccctcatcttttaggggaatcattaccaggtttggagaggattccagagcgtcag 120
QY 2543 tgctttcaataatgtctcgaactctgcctctttgttctcattgattgattcgaataaa 2602
Db 121 tgctttcaataatgtctcgaactctgcctctt--tgctcattgattgattcgaataaa 178
```


Search completed: July 31, 2001, 07:27:42
Job time: 17874 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 08:15:33 ; Search time 10902.7 Seconds
(without alignments)
5081.817 Million cell updates/sec

Title: US-09-402-713A-6
Perfect score: 3582
Sequence: 1 acagaagaataagcaagtgc.....tgattctttgttacaacttt 3582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_ba3:
4: gb_in1:
5: gb_in2:
6: gb_in3:
7: gb_om:
8: gb_ov:
9: gb_pat1:
10: gb_pat2:
11: gb_ph:
12: gb_pl1:
13: gb_pl2:
14: gb_pl3:
15: gb_pl4:
16: em_ba1:
17: em_ba2:
18: em_fun:
19: em_htgo_hum:
20: em_htgo_inv:
21: em_htgo_rod:
22: em_htg_hum1:
23: em_htg_hum2:
24: em_htg_hum3:
25: em_htg_hum4:
26: em_htg_hum5:
27: em_htg_hum6:
28: em_htg_hum7:
29: em_htg_hum8:
30: em_htg_inv1:
31: em_htg_inv2:
32: em_htg_other:
33: em_htg_rod:
34: em_hum1:
35: em_hum2:
36: em_hum3:
37: em_hum4:
38: em_hum5:
39: em_hum6:
40: em_hum7:
41: em_in:
42: em_om:
43: em_or:

44: em_ov:
45: em_pat:
46: em_ph:
47: em_pl:
48: em_ro:
49: em_sts:
50: em_sy:
51: em_un:
52: em_vi:
53: gb_sts1:
54: gb_sts2:
55: gb_sts3:
56: gb_sy:
57: gb_un:
58: gb_v1:
59: gb_v2:
60: gb_htg1:
61: gb_htg2:
62: gb_htg3:
63: gb_htg4:
64: gb_htg5:
65: gb_htg6:
66: gb_htg7:
67: gb_htg8:
68: gb_htg9:
69: gb_htg10:
70: gb_htg11:
71: gb_htg12:
72: gb_htg13:
73: gb_htg14:
74: gb_htg15:
75: gb_htg16:
76: gb_htg17:
77: gb_htg18:
78: gb_htg19:
79: gb_htg20:
80: gb_htg21:
81: gb_htg22:
82: gb_htg23:
83: gb_htg24:
84: gb_htg25:
85: gb_pr1:
86: gb_pr2:
87: gb_pr3:
88: gb_pr4:
89: gb_pr5:
90: gb_pr6:
91: gb_pr7:
92: gb_pr8:
93: gb_pr9:
94: gb_ro1:
95: gb_ro2:
96: gb_in4:
97: gb_pr10:
98: em_ba3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	3497	97.6	3923	88	AF103907	AF103907 Homo sapi
2	3032.4	84.7	5435	88	AF103908	AF103908 Homo sapi
3	2959.8	82.6	164371	80	AL390239	AL390239 Homo sapi
4	1363.4	38.1	173831	80	AL359314	AL359314 Homo sapi
c 5	1319.8	36.8	267581	80	AL358573	AL358573 Homo sapi
c 6	536	15.0	546	54	G56925	G56925 SHGC-102485
c 7	533	14.9	542	54	G56926	G56926 SHGC-102486
8	235.4	6.6	359	9	AX018075	AX018075 Sequence

```
9 121 3.4 143675 79 AL161625 Homo sapi
10 121 3.4 267581 80 AL358573 Homo sapi
11 120 3.4 580 89 AF279290 Homo sapi
C 12 119.4 3.3 172298 61 AC009556 Homo sapi
13 77.4 2.2 7218 10 I66494 Sequence 14
C 14 57.4 1.6 7218 10 I66494 Sequence 14
C 15 56 1.6 163266 70 AC027016
C 16 53.8 1.5 502 7 BTMISATP
17 53 1.5 1141 10 AX083744 Sequence
18 52.2 1.5 86064 81 AL391554 Homo sapi
C 19 52.2 1.5 143489 79 AL161637 Homo sapi
C 20 51.6 1.4 182999 63 AC013260 Homo sapi
21 50.8 1.4 161222 64 AC016037 Homo sapi
22 50.8 1.4 169942 80 AL365230 Homo sapi
C 23 50.4 1.4 118981 71 AC040983 Homo sapi
24 50.2 1.4 153054 81 AL450423 Homo sapi
25 50 1.4 200000 60 AC007705 Homo sapi
C 26 49.2 1.4 199563 82 AL590503 Mus muscu
C 27 48.8 1.4 291 53 AU025050 Rattus no
C 28 48.8 1.4 201885 66 AC020703 Homo sapi
29 48.6 1.4 112902 85 AC005230 Homo sapi
30 48.6 1.4 145554 84 CNS07BEU Homo sapi
31 48.4 1.4 155103 76 AC083837 Homo sapi
32 48.4 1.4 167075 92 HS742324 Human DNA
C 33 48.4 1.4 186572 75 AC079130 Mus muscu
C 34 48.2 1.3 113367 86 AC008178 Homo sapi
C 35 48 1.3 157676 70 AC027358 Homo sapi
C 36 48 1.3 180392 65 AC019299 Homo sapi
C 37 48 1.3 205053 87 AC011749 Homo sapi
C 38 47.8 1.3 66863 71 AC036112 Homo sapi
C 39 47.4 1.3 74266 77 AC087104 Arabidops
40 47.4 1.3 80376 12 AB010073 Arabidops
41 47.4 1.3 130705 92 HS232122 Homo sapi
C 42 47.2 1.3 1141 10 AX083744 Sequence
C 43 47.2 1.3 137064 74 AC073363 Homo sapi
C 44 47.2 1.3 152371 66 AC021589 Homo sapi
C 45 47.2 1.3 160133 69 AC025140 Homo sapi
```

ALIGNMENTS

```
RESULT 1
AF103907 3923 bp mRNA 14-AUG-2000
LOCUS Homo sapiens non-coding RNA DD3 sequence.
ACCESSION AF103907
VERSION AF103907.1 GI:6165973
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3923)
AUTHORS Bussemakers,M.J., van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F., Schalken,J.A., Debruyne,F.M., Ru,N. and Isaacs,W.B.
TITLE DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072260
PUBMED 10606244
REFERENCE 2 (bases 1 to 3923)
AUTHORS Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru,N. and
Isaacs,W.B.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
Location/Qualifiers
1..3923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
```

|||||
Db 841 ATATTTCTAACCCCTAAAACAAAGCTGTTGTAATATCTGATCTCTACGGTTCCTCTCGG 900
QY 901 cccaacattctccatatcaagccacacactcaatttttaattattagttccceagatctga 960
Db 901 CCCACATTTCCATATATCCAGCCACACTCATTTTAAATATTTAGTTCGCCAGATCTGTA 960
QY 961 ctgtgacctttctacactgtagaataaacattactcattttgttccaagacccttcgtgt 1020
Db 961 CTGTGACCTTTTACACTGTAGATAAACAATTAATCATTTGTTCAAAGACCCCTCGTGT 1020
QY 1021 gctgcttaataatgtagctgactgtttttcttaagagtggtcttgccccaggggatctgtg 1080
Db 1021 GCTGCCATAATATGTAGCTGACTGTTTTTCTTAAGGAGTGTCTGCCCCAGGGGATCTGTG 1080
QY 1081 aacaggctgggaagcatctcaagatctttccagggtttactactactagcacacagcatga 1140
Db 1081 AACAGGCTGGGAAGCATCTCAAGATCTTTTCCAGGGTTATACTTACTAGCACACAGCATGA 1140
QY 1141 tcattacggagtgaattatcttaataacatcatctccagtgctctttggccatactagaat 1200
Db 1141 TCATTTACGGAGTGAAATATCTAATCAACATCATCTCAGTGCTTTGGCCCATACTGAAT 1200
QY 1201 tcaattcccaactttgtgccccattctcaagacctcaaaaaatgtaattccatttaataatca 1260
Db 1201 TCATTTCCCACTTTTGTGCCATTTCTCAAGAGCTCAAAATGTCATTTCCATTAATATACA 1260
QY 1261 gattataacttttttttaacctggaagaattcaattcaatgttacatgcagctatgggaattta 1320
Db 1261 GGATTAACTTTTTTTTTAAACCTGGAGAAATTCATTTGTTACATGACAGCTATGGGAATTTA 1320
QY 1321 attacaattttgtttccagtgcaagatgactaagctccttcttccctccctttgttt 1380
Db 1321 ATTACATATTTGTTTCCAGTGCAAGATGACTAAGTCTCTTATCCCTCCCTTTGTTT 1380
QY 1381 gattttttccagtaataagttaaaaatgcttagccttgtagctgtgtagggctgtatcacagca 1440
Db 1381 GATTTTTTTTCCAGTATAAAGTTAAATGCTTAGCCTTGTACTGAGCGTGTATACAGCAC 1440
QY 1441 agcctctcccaactccctcagcttatctgtcatacaccatcaacccctccctaccacct 1500
Db 1441 AGCCTCTCCCACTCCCTCCAGCCTTATCTGTATACCACTCAACCCCTCCCACTACCACCT 1500
QY 1501 aacaaaaactaacttgtaattccttgaacatgtccaggaacatacatattcctcttgct 1560
Db 1501 AAACAAAATCTAATCTGTAATCTTGAACATGTCAGGACATACATATTCTCTGCT 1560
QY 1561 gagaagctcttcctgtctcttaaatcttagaatgtatgaagtgttgaagtgttagacta 1620
Db 1561 GAGAAGCTCTTCTCTCTTAAATCTAGAAATGATGATGAAGTTTGTGAATTAAGTTGACTA 1620
QY 1621 tcttaactcatgaagaaggagcacatatgagattcatcatcacatgagacagcaata 1680
Db 1621 TCTTACTTCATGCAAGAAGGACACATATGATGATTCATCATCATGAGACAGCAATA 1680
QY 1681 ctaaaagtgtaattgtattagaagtttagataaataatgaaatgcaagaccacaga 1740
Db 1681 CTAAAAAGTGAATTTGATTTATAGAGTTTAGATAAATATATGAAATGCAAGGCCACAGA 1740
QY 1741 gggaaatgtttatggggcagctttgtgaagcctgggatgtgaaagcaaggcaggaaacctca 1800
Db 1741 GGGAAATGTTTATGGGCGACCTTGTGAAGCCTGGGATGTGAAGCAAAAGGCGAGGAACTCA 1800
QY 1801 tagtatcttataataatacttcattctctatctctatcaacaataatcaacaagctttt 1860
Db 1801 TAGTATCTTATATAATATACTTCAATTTCTATCTCTATCACAATATCCAACAAGCTTTT 1860
QY 1861 cacagaaticatgcagtgcaaatccccaaaggtaacctttatccatttcaatgtgtagtg 1920
Db 1861 CACAGAAATTCATGCGAGTGCAAAATCCCCAAAGGTAACCTTTATCCATTTTCATGGTAGTGC 1920
QY 1921 gctttagaatatttggcaaatcatactggtgcaacttatctcaactttgagatgtgtttgtcc 1980
|||||

Db 1921 GCTTTAGAAATTTTGGCAATCATACTGCTCACTTATCTCAACTTTGAGATGTGTTGTCC 1980
QY 1981 ttgtagtttaattgaagaataatagcgacactctgtgaagcacttttaggtttcactcctggc 2040
Db 1981 TTGTAGTTAATTTGAAGAANAATAGGCGACACTTTGTGAGCCACTTTTAGGGTTCACCTCTGGC 2040
QY 2041 aataaagaattttacaagaagctactcaggaccaggtttgtaagagctctgtgtgtgtgt 2100
Db 2041 AATAAAGAATTTTACAAAGAGCTACTCAGGACCAGTGTGTTAAGAGCTCTGTGTGTGTGT 2100
QY 2101 ggtgtgtgtgagttacatgccaagaagtgtgcctctctctctctgtgacctattttcagac 2160
Db 2101 GTGTGTGTGTGAGTGTACATGCCAAAGTGTGCTCTCTCTCTGTGACCCATTTATTTTCAGAC 2160
QY 2161 ttaaaacagcatgttttcaaatggcactatgactgccaatgactgccaatgactatccacacat 2220
Db 2161 TTAAACAAGCATGTTTTCAATGGCACHATGAGCTGCCAATGATGATATCACACCATAT 2220
QY 2221 ctcatattctccagtaaatgtgataaataatgtcatctgttaacataaaaaaagttttgac 2280
Db 2221 CTCATTTATTTCCAGTAAATGTGATAAATAATGTCACTCTGTTAACATAAAAAAAGTTTGAC 2280
QY 2281 ttcaaaaaagcagctggaaatggacaacacaaataatgataaataatcttaactcctaccatca 2340
Db 2281 TTTCACAAAAGCAGCTGGAATGGACAAACCAATATGTCATAAATCTTAACCTCTACCATCA 2340
QY 2341 gctcacactgcttgacatatattgttagaagcacctcgactttgtggttctctcttaagc 2400
Db 2341 GCTACACACTGCTTGACATATATGTTTAGAACACCCTCCATTTGTGGGTTCCTCTTAAGC 2400
QY 2401 aaaaacttctcagctagctcagctgggctgtgcatcagcggttttgagaaataattcaa 2460
Db 2401 AAAATACTTTGCATTTAGTCTCAGCTGGGCTGTGTCATCAGCGGTTTGTAGAAATAATTCAA 2460
QY 2461 ttctcagaagaagcagaatttgaattccctcatctcttttaggaatacattaccaggtttg 2520
Db 2461 TTCTCAGCAAGAACCCAGAAATTTGAATTCCTCATCTCTTTTAGGAATCATTTACCAGGTTTG 2520
QY 2521 gagagattcagacagctcaggtgctttcactaatgtctctgaaactctgtccctctttg 2580
Db 2521 GAGAGATTTACACAGCTCAGGTGCTTTTCACTAATGTCTCTGNACTTCTGCTCCCTTTTG 2580
QY 2581 tgttcagtgatagttccaaataaataatgttatctttgaactgatgctcatagagagaata 2640
Db 2581 TGTTCATGATAGTCCCAATAAATAATGTTATCTTTTGAACCTGATGCTCATAGGAGAGAATA 2640
QY 2641 taagaactctgagtgatatacaacattagggattccaagaataattagatttaagctccaca 2700
Db 2641 TAAGAACTCTGAGTGATATCAACATTTAGGGATTTCAAAGAAATATTAGATTTTAAGCTCACA 2700
QY 2701 ctggtcaaaagaagcaagatacaagaactctgagctgtcatcgccccatctctgtga 2760
Db 2701 CTGGTCAAAAGGAACCAAGATACAAAGAACTGTGAGCTGTCTATCGTCCCATCTCTGTGA 2760
QY 2761 gccacaaccaacagcagggaccacacgcagctgtctgagatcctttaaatacaaggaaacagtg 2820
Db 2761 GCCACAACCAACAGCAGGACCCCAACCGCTGTGAGATCTTTAAATCAAGGAACCAAGTG 2820
QY 2821 tcatgagttgaattctctctattatgagtagctctgtgacctctgacctctctgctctctct 2880
Db 2821 TCATGAGTTGAATTTCTCTCTATTATGGATGCTAGTCTTGTGCCATCTCTGCTGCTCTCT 2880
QY 2881 gacacatatagcttctagcctttgtcttcacagacttttatctttctccacacacacac 2940
Db 2881 GACACATATTTAGCTTTCAGCTTTTGTCTTCAGACTTTTATCTTTTCTCCACACATCGC 2940
QY 2941 ttaccaactctctctctgtctgtgtgtcttttgaccttccccacaaagaatttcaacgactct 3000
Db 2941 TTACCAATCTCTCTCTGCTGTCTGCTTTGGACTTCCCCACAAGAAATTTCAAGACTCT 3000
QY 3001 caagtcttttcttccatccccaccactaacctgaattgctgtagaccttatcttttattaa 3060
Db 3001 CAAGTCTTTTCTTCCATCCCCACCACCTAACCTGAA-TGCCTAGACCCCTTATTTTTATTAA 3059

```

QY 3061 ttccaatagatgctgcctatgggctaataattgcttttagatgaacattagatatttaaa 3120
Db 3060 TTTCCAATAGATGCTGCTATGGCT-ATATTGCTTTAGATGAACATTAGATATTTAAAG 3118
QY 3121 tctaaagaggttcaaaatccaactcaattatcttcttcttcttccactccctcctct 3180
Db 3119 CTCGAAGAGGTTCAAAATCCAAATCAATATCTTCTTCTTCTTCCACCT-CCCTGCTCCTCT 3177
QY 3181 cctatatattactatg-actgaacaggatggtcccaaa-gatgccagtcacaaatgagaaa 3238
Db 3178 CCTATATTAATGATGTCATGAACAGCATGGTCCCAATGTAGCCATGCAAAATGAGAAA 3237
QY 3239 cccagtggtcccttggtgatcatgcatgcaagactgctgaagccag-aggatgactgatt 3297
Db 3238 CCCAGTGCTCCTTGTTGGTACATGCATGCAAGACTGCTGAAGCCAGAGGATGACTGATT 3297
QY 3298 acgcctcatgggtggaggggaccactcctcctggcctctgtagtgcagagcaagacctg 3357
Db 3298 ACGCCCTCATGGGTGGAGGGGACCACCTCCTGGGCTTTCGTGATTGTGAGGAGCAAGACCTG 3357
QY 3358 agatgctccctgccttcagtgctcctgcatctcccttcttcttaatgaagatccatagaat 3417
Db 3358 AGATGCTCCTGCTTCAGTGTCTCTGCACTCCCTTCTTAATGAAGATCCATAGAAAT 3417
QY 3418 ttgctacatttgagaattccaattaggaactcacatgctgttttctctgcctatcaatttt 3477
Db 3418 TTGCTACATTTGAGAAATCCAAATAGGAACACATGTTTATCTGCCCTACATTTT 3477
QY 3478 taaactgctgaaataaagtcttttcaaaaatcgtcctgtgtaatacttttcttaca 3537
Db 3478 TAAACTGCTGAAATTAAGTTTTTCAAAATCTGCTTGTAAATTAATCTTTTCTTACA 3537
QY 3538 gtcttgccatactatatacaacttgattcttctgtttacaaattt 3582
Db 3538 GTGCTTGCCATACTATATCAACTTTGATTCTTTGTACAACTTT 3582

RESULT 2
AF103908 5435 bp DNA PRI 14-AUG-2000
LOCUS Homo sapiens non-coding RNA DD3 gene, exons 2, 3, and 4.
ACCESSION AF103908
VERSION AF103908.1 GI:6165974
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5435)
AUTHORS Bussemakers,M.J., van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaush,H.F., Schalken,J.A., Debruyne,F.M., Ru,N. and Isaacs,W.B.
DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072260
PUBMED 10606244
REFERENCE 2 (bases 1 to 5435)
AUTHORS Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaush,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru,N. and
Isaacs,W.B.
Direct Submission
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
Source
1..5435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q21-q22"
repeat_region 1..78
/rpt_family="AluY"
/rpt_type=dispersed

```

```

exon 533..697
/note="alternative exon present in 5% of cDNA clones"
/number=2
repeat_region 1035..1294
/rpt_family="Alu"
/rpt_type=dispersed
mRNA join(1571..1753,1981..5435)
/note="transcript III"
mRNA /product="non-coding RNA DD3"
join(1571..1753,1981..2517)
/note="transcript I"
mRNA /product="non-coding RNA DD3"
join(1571..1753,1981..3579)
/note="transcript (major) II"
exon /product="non-coding RNA DD3"
1571..1753
/number=3
exon 1981..5435
/number=4
polyA_signal 2495..2499
/note="transcript I"
polyA_site 2517
/note="transcript I"
polyA_signal 3553..3558
/note="transcript (major) II"
polyA_site 3579
/note="transcript (major) II"
repeat_region 5268..5423
/note="LINE"
polyA_site 5435
/note="transcript III"
BASE COUNT 1546 a 1188 c 1036 g 1654 t 11 others
ORIGIN

Query Match 84.7%; Score 3032.4; DB 88; Length 5435;
Best Local Similarity 99.5%; Pred No. 0;
Matches 3105; Conservative 0; Mismatches 11; Indels 6; Gaps 6;

QY 464 cagagatccctgggagaaatgccggccgccatcttgggtcatcgatgagctcgcctg 523
Db 1976 CACAGATCCCTGGGAGAAATGCCGGCCGCCATCTTGGGTCAATCGATGAGCTCGCCCTG 2035

QY 524 tcctggtccctgtgtggaagacattagaaaatgaattgattgttctctaaag 593
Db 2036 TGCTGTGCTCCCTGTGTGAGGAAGACATTAGAAATGAATTGATGTGTCTCTTAAAGG 2095

QY 584 atgggcagaaacagatcctgtgtggatattttgaacgggattcacagatttgaa 643
Db 2096 ATGGGCAGAAACAGATCCTGTGTGGATATTTATTGAACGGGATTACAGATTTGAAA 2155

QY 644 tgaagtcaaaagtgcattaccatgagagaaacagagaaaaatcttgatgct 703
Db 2156 TGAAGTCAAAAGTGAGCATTTACCAATGAGAGAAACAGACGAGAAAAATCTTTGATGGCT 2215

QY 704 tcacaagacatgcaacaaacaaatggaatactgtgatgacatgagcgagcaagctgg 763
Db 2216 TCACAAGACATGCAACAAACAAATGGAATGGAATGATGTGATGATGAGCGACGCAAGCTGGG 2275

QY 764 gaggagataaccacggggcagaggggtcaggattcttgccctgctgcctgctgctt 823
Db 2276 GAGGAGATAACACAGGGGCGAGGGGTCAGGATTTCTGCCCTGCTGCTTAACTGTGCGTT 2335

QY 824 cataaccataatcatttcatttcttaacctcaaaacaaagctgttgaatatctgatct 883
Db 2336 CATAACCAAAATCATTTTCATATTTCTAACCTTCAAAACAAAGCTGTTGTATATCTGATCT 2395

QY 884 ctacggttctctggggcccaacattctccatatatccagccacacactatcttatt 943
Db 2396 CTACGGTTCCTTCTGGGCCCCAACATTTCTCCATATATATCCAGCCACACTCATTTTAAATTT 2455

```


Db	4614	ACATTAGATATTTAAAGCTCAAGAGGTTCAAAATCCAACTCATATATCTCTCTTTCTTTC	4673
Qy	3164	acctcccctgctcctccctatactactgattg-actgaacagagatgggtcccaaa-gat	3221
Db	4674	ACCT-CCCTGCTCCTCTCCCTATATTACTGATTCGACTGAACAGCATGGTCCCCAATGTA	4732
Qy	3222	gcaagtcaaatgagaacccaagtggtcctctgttgatgcatgcatgaagactgctgaagc	3281
Db	4733	GCCATGCAAAATGAGAAACCCAGTGGCTCCCTGTGTGGTACATGTCATGCAAGACTGCTGAAGC	4792
Qy	3282	cag-aggatgactgattacgcctcatgggtggagggagccacccctggtggccttcggtatt	3340
Db	4793	CAGAAGATGACTGATTACGCCTCATGGGTGGAGGGGACCACCTCCCTGGGCTTCGTGATT	4852
Qy	3341	gtcaggagcaagacctgagatgctccctgccttcagtgctcctctgcatcctccctctcta	3400
Db	4853	GTCAGGAGCAAGACCTGAGATGCTCCCTGCCTTCAGTGTCTCTGTCGATCCTCCCTTTCTA	4912
Qy	3401	atgaagatccatagaattgctacatttgagaattccaattaggaaactcacatgttttat	3460
Db	4913	ATGAAGATCCATAGAAATTGCTACATTTGAGAATTCCAATTAGGAACTCACATGTTTAT	4972
Qy	3461	ctgccctatcaattttttaaactgtcgaataaataagtttttcaaaaatcgtcctctgta	3520
Db	4973	CTGCCCTATCAATTTTTTAACTTGCTGAAATTAAGTTTTTTCAAAATCTGCTCTGTA	5032
Qy	3521	aattactttttcacagtctctggcatactatatcaacttgatctcttctgttacaact	3580
Db	5033	AATTACTTTTTCTTACAGTGTCTTGGCATCTACTATCAACTTTGATTCCTTTGTTCACAACT	5092
Qy	3581	tt 3582	
Db	5093	TT 5094	
RESULT 3			
AL390239			
LOCUS	AL390239	164371 bp	DNA
DEFINITION	Homo sapiens chromosome 9 clone RP11-58J3, *** SEQUENCING IN		
ACCESSION	AL390239		
VERSION	AL390239.11		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 164371)		
JOURNAL	Burton, J.		
	Direct Submission		
	Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,		
	CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk		
	requests: clonerequests@sanger.ac.uk		
COMMENT	On Mar 12, 2001 this sequence version replaced gi:13273805.		
	----- Genome Center		
	Center: Sanger Centre		
	Center code: SC		
	Web site: http://www.sanger.ac.uk		
	Contact: humquery@sanger.ac.uk		
	----- Project Information		
	Center project name: BA58J3		
	----- Summary Statistics		
	Assembly program: XGAP4; version 4.5		
	Sequencing vector: Plasmid; L08752; 100% of reads		
	Chemistry: Dye-terminator Big Dye; 100% of reads		
	Consensus quality: 163337 bases at least Q40		
	Consensus quality: 163481 bases at least Q30		
	Consensus quality: 163577 bases at least Q20		
	Insert size: 164071; sum-of-contigs		
	Insert size: 166918; 4.2% error; agarose-fp		
	Quality coverage: 8.78x in Q20 bases; sum-of-contigs Quality		
	coverage: 9.07x in Q20 bases; agarose-fp		

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 20240: contig of 20240 bp in length
 * 20241 20340: gap of 100 bp
 * 20341 116100: contig of 95760 bp in length
 * 116101 116200: gap of 100 bp
 * 116201 118243: contig of 2043 bp in length
 * 118244 118343: gap of 100 bp
 * 118344 164371: contig of 46028 bp in length.

FEATURES
 source
 1..164371
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-58J3"
 /clone_lib="RPC1-11.1"
 1..20240
 /note="assembly_fragment:00838
 fragment_chain:1
 clone_end:SP6
 vector_side:left"
 20341..116100
 /note="assembly_fragment:01998
 fragment_chain:1"
 116201..118243
 /note="assembly_fragment:00916
 fragment_chain:1"
 118344..164371
 /note="assembly_fragment:00334
 fragment_chain:1
 clone_end:T7
 vector_side:right"

BASE COUNT 50254 a 34905 c 32663 g 46249 t 300 others
 ORIGIN

Query Match 82.6%; Score 2959.8; DB 80; Length 164371;
 Best Local Similarity 99.2%; Pred. No. 0;
 Matches 3100; Conservative 0; Mismatches 12; Indels 13; Gaps

QY 454 cagagatccctgggagaatgccgcggccgaccttctgggtcatcgatgagcctcgccctg 523
 Db 33316 CACAGATCCCTGGGAGAAATGCCGCGCCCATCTTGGGTCAATGATGTTCTTTAAAGG 33375

QY 524 tgcctgtccgcgttgtagggagaagacattagaaaatgaattgatgtgttccttaaaagg 583
 Db 33376 TGCCCTGGTCCCGCTTGTGAGGGAAGGACATTAGAAAATCAATTGATGTGTTCTTTAAAGG 33435

QY 584 atgggcagaaaaacagatcctgtgtgatatatttgaacgggattacagatttgaaa 643
 Db 33436 ATGGCAGGAAAAACAGATCCCTGTTGGGATATTTATTTGAACGGGATTACAGATTTGAAA 33495

QY 644 tgaagtcaaaagtgcgacttaccaatgagaggaaaacagacgaaaaatcttctgatgct 703
 Db 33496 TGAAGTCACAAAGTGAGCATTTACCAATCAGAGGAAAACAGAGCAAAATCTTGATGGCT 33555

QY 704 tcacaagacatgcacaaacaaaaatggaaatactgtgatgacatgaggcagccaagctggg 763
 Db 33556 TCACAAGACATGCAACAACAACAAAATGGAAATACCTGTGTGATGACATGAGGAGCAAGCTGGG 33615

QY 764 gagagataaacacaggggcagaggctcaggattcttgccctgctcctctaaactgtgcatt 823
 Db 33616 GAGGAGATTAACACGCGGCAGAGGCTCAGGATCTTGCGCCCTGCTGCCCTAAACTGTGCGTT 33675

QY 824 cataaccaaatcattctcatattcttaaccctcaaaacaaagctgttgtaatactgatct 883
 Db 33676 CATACCAAAATCATTTCTATATTTCTAACCCCTCAAAACAAAGCTGTGTAATATCTGACTCT 33735

Qy	884	ctacggttcctctcgtggcccaaacattctccatatataccagccacactcatttttbaatatt	943
Db	33736	CTACGGTTCTCTCTGGGCCCAACATCTCCATATATCCAGCCACACTCATTTTAAATATT	33795
Qy	944	tagttccagatctgactgtgaccttctacacgttagaataacattactcatttggtt	1003
Db	33796	TAGTTCACGATCTGTACTGTGACTCTTCTACACTGTAGAAATAACATTACTTCATTTTGGTT	33855
Qy	1004	caaaagacccttcgttgtctgcctaaatagttagctgactgtttttcctaaaggagtctct	1063
Db	33856	CAAAGACCCTTCCGTGTGCTGCCTTAATATGTAGCTGACTGTTTTCCTTAAGAGFGTCT	33915
Qy	1064	ggcccaagggaatctgttgaaacaggctgggaagcatctcaagaatcttccaggggttatactt	1123
Db	33916	GGCCCGAGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTTCAGGGTTATACTTT	33975
Qy	1124	actagcacacagcatgatcattacggagtgaaattactaaaccaatcatcctcagtgctc	1183
Db	33976	ACTAGCACACAGCATGATCATTACGGAGTGAATTTATCTAATCAACATCATCCTCAGTGCTC	34035
Qy	1184	tttgcccatactgaaattcatttcccacttttggcccattctcaagacctcaaaaatgctc	1243
Db	34036	TTTGCCCATACTGAAATTCATTTCCCACTTTTGTGCCATTTCTCAAGACCTCAAAATGTC	34095
Qy	1244	attccattaatatcacaggattaaacttttttttaaccctgggaagaattcaatgtttacat	1303
Db	34096	ATTCCATTAAATATCACAGGATTAACCTTTTTTTTAACTTGGGAAGATTCAAATGTTACAT	34155
Qy	1304	gcagctatgggaatttaattacatatttggtttttccagtgcaaaagatgaactcacttt	1363
Db	34156	CGAGCTATGGGAATTAATTACATATTGTGTTTCCAGTGCAAGAGTACACTAAGTCCTTT	34215
Qy	1364	atccctcccctttgttgatttttttccagataaaagttaaaaatgcttagccttgact	1423
Db	34216	ATCCCTCCCCTTTGTTTGATTTTTTTTTTCCAGTATAAGTTAAAAATGCTTAGCCCTTGACT	34275
Qy	1424	gagcgtatcacag-cacagcctctcccactcccctccagccttatctgtcatacaccatca	1482
Db	34276	GAGGCTGTATACAGCCACAGGCTCTCCCCATCCCTCCAGCCCTTATCTGTCTATCACCATCA	34335
Qy	1483	accctcccctaccacactaaacaaaactaacttgtaattccttggacatgtcaggacat	1542
Db	34336	ACCCTCCCAT - GCACCTAAACAAAATTAACCTTGTAATTCCTTGAACATGTCAGG - CAT	34393
Qy	1543	acattattcctctgcctggagaagctctcctgtctctaaactcagaatgatgataaag	1602
Db	34394	ACATTTATCTCTCTGGCTGAGAAGCTCTCCCTGTCTCTTAAATCTAGAATGATGTAAAG	34453
Qy	1603	tttgaataagttgactatcttacttcatgcaaaagaggacacatatgagatcatcat	1662
Db	34454	TTTTGAATAAGTTGACTATCTTACTTTCATGCAAGAGGGACACATATGAGATTCATCAT	34513
Qy	1663	cacatgacagcaaaatactaaaaagtgtaattgattataaagagtttagataaaatatag	1722
Db	34514	CACATGACAGCAAAATACTAAAAAGTGTAAPTTCATTATAAGAGTTTTAGATAAATAATAG	34573
Qy	1723	aaatgcaagagcacacagaggaatggtttatgggcacgtttgtaagcctggagtgtaag	1782
Db	34574	AAATGCAAGAGCCACAGAGGAATGTTTATGGGGCAGCTTTGTAAAGCTGGGATGTGAAG	34633
Qy	1783	caaaaggcaggaaacctcatagtaacttataataataacttacttctctactctactaac	1842
Db	34634	CAAAGCAGGGAACCTCATAGTATCTTTATATAATATACTTCATTTCTTATCTCTATCAC	34693
Qy	1843	aataaccaaaagcttttcacagaattcatgctagctgcaaatcccccaaggtaaacctttat	1902
Db	34694	AATATCCAACAAGCTTTTTCACAGAAATTCATGCACTGCGAAATCCCCCAAGGTAAACCTTTAT	34753
Qy	1903	ccatttcaggtgagtcgcttttagaatttttgcaaatcatactggttcaacttatctcaac	1962
Db	34754	CCATTTTCATGCTGAGTCGGCTTTTGAATTTTGGCAATATACACTGCTACCTTATCTCAAC	34813

	Qy	1963	t t g a g a t g t g t t g t c c t t g t a g t t a a t g a a g a a a t a g g c a c t c t t g t a g c c a c t	2022
	Db	34814	T T T G A G A T C T G T T T T G C T C T T G T A G T T A A T T G A A G A A A T A G G C A C T C T T G T G A G C C A C T	34873
	Qy	2023	t t a g g t t c a c t c c t g c c a a t a a g a a t t t c a a a g a c t a c t c a g a c c a g t t g t a a g	2082
	Db	34874	T T A G G G T T C A C T C C T G G C A A T A A A G A Y T T C A A A G A G C T A C T C A G G A C C A G T T G T T A A G	34933
	Qy	2083	a g c t c t g t g t g t g t g t g t g t g a g t g a c t a g c c a a a g t g t g c c t c t c t c -	2141
	Db	34934	A G C T C - ~ G T G T G T G T G T G T G T G A G T G A C A T C C A A A G T G C C T C T C T C T	34991
	Qy	2142	t t g a c c c a t t t t c a g a c t t - a a a c a a g c a t g t t t t c a a t g g c a c t a g a g t g c c a	2200
	Db	34992	T T G A C C C A T A T T T C A G A C T T A A A A A C A A G A N T G T T T T C A A T G G C A C A T A T A G A G T G C C A	35051
	Qy	2201	a t g a t g a t c a c c a c a t a t c a t t a t t c c a g t a a a t g t a a a t g t a a a t a a t g t a c a t c g t	2260
	Db	35052	A F A G T A T A C C A C C A T A T C A T A T T C C A G A A T G A T A A T T A A T G T C A T C T G T	35111
	Qy	2261	L a a c a t a a a a a a g t t t g a c t t c a a a a g c a g t g g a a t g g a c c a c c a c a a t a t g e a t	2320
	Db	35112	T A A C A T A A A A A A G T T G A C T T C A C A A A A G C A G T G G A A T G G A C A C C A C A N A T A T G A T	35171
	Qy	2321	a a a t c a a c t o c t a c c a t c a g t a c a c a c t g o t t g a c a t a t a t t g t t a g a a g c a c c t g c y	2380
	Db	35172	N A A T T A A C T C T A C A C A C A G C A C A C A C T G C T T G A C A T A T A T T T A G A A G A C A C C T G C	35231
	Qy	2381	a t t t g a g t t c t c t a a g a a a a c t g t a a t g a g t c a c t c a g t g g g t g t g c a t c a g	2440
	Db	35232	A I T T G T G G G T C T T T A A G C A A A A T A C T T G C A T A G G T C T C A G C T G G G G C T G T G C A T C A G	35291
	Qy	2441	c g g g t t t g a g a a a t t c a a t t c t c a g a a g c c a g a a t t t g a a t c c c t c a t c t t t t a	2500
	Db	35292	G O G G T T T G A G A A A T A T T C A A T T C T C A G C A A G C C A G A A T T T G A A T T C C C T C A C T T T T A	35351
	Qy	2501	g g a a t c a t t a c c a g t t t g a g a g a t c a g a c a g c t c a g t g o t t b c a c t a a t g t c t c	2560
	Db	35352	G G A A P C A T T T A C C A G T T T G G A G A G A T T C A G A C A G C T C A G T G T T T C A C T A A T A T G T C	35411
	Qy	2561	I q a a c t c t g l c c c t c t t g t c a t g a t a g t c c a a t a a a t g t t a c t t g t a a c t	2620
	Db	35412	T G A A C T T C T G C C C T C T T T G T G T C A T G G A T A G T C C A A T A A T A T G T T A T C T T T G A A C T	35471
	Qy	2621	g a t g t c a t a g g a g a a t a L a a g a a c t c t g a t g a t a c a a t a g g a t t a g g a t c c a a g a a	2680
	Db	35472	G A T G C T C A T A G A G A A T A T A A G A A C T C T G A T G A T A T A A C A T T A G G G A T T C A A A G A A	35531
	Qy	2681	a t a t a g a t t a a g t c a c a c t g t t c a a a a g g a c c a a g a t a c a a g a a c t c t g a g t g t	2740
	Db	35532	A T A T T T A G A T T T A A G T C A C A C T G G T C A A A A G G A A C C A A G A T A C A A A A A C T C T G A G T G T	35591
	Qy	2741	C a t g t c c c a t c t c t g t a g c c a c a c c a a c a g c a g a c c a a c a g a t g t c t a g a t c c	2800
	Db	35592	C A T G T C C C A F C T C T G T A G C C A C A C C A C A G C A G G A C C C A A C G C A T G T C T G A G A T C C	35651
	Qy	2801	t t a a a t c a a g a a a c c a g t c a t g a t t g a a t t c t c t a t t a t g a t g c t a g c t c t g g	2860
	Db	35652	T T A A A T C A A G A A A C C A G T G C A T G A G T T G A A T T C T C T A T A T A T G A T G C T A G C T T C T G G	35711
	Qy	2861	O a t c t c t g g t c t c t c t t g a c a c a t a t a g c t t c a g c t t g c t t g c c a c a g a c t t t t a	2920
	Db	35712	C C A T C T C T G G C T C T C C T T T G A C A C A T A T A G C T T C T A G C T T T G C T T C C A C A G A C T T T T A	35771
	Qy	2921	t c t t t t c c a a c a c a t c g e t t a c c a a t c c t c t c t c t g t c t g t g t t g t t g a a c t t c c c c	2980
	Db	35772	T C T T T T T C C A A C A C A T C G C T T A C A A T C C T C T C T C T G T T G C T T T G S A C T T C C C C	35831
	Qy	2981	a c a a g a a t t c a a g a c t c t c a a g t c t t t t t c t c a c c c a c c a c t a a c c t g a a t g c c	3040
	Db	35832	A C A A G A A T T T C A A C G A C T C T C A A G T C T T T T C T T C C A T C C C C A C C A C T A A C C T G A A - T G C C	35890
	Qy	3041	L a g a c c c t a t t t t t a a t t t c c a a t a g a t g t c c t a t g g g c t a a t a t g t g t t a g a	3100

```
Db 35891 TAGACCTTATTTTATTAATTTCCCAATAGATGCTGCTATGGCT-ATATGCTTTAGA 35949
QY 3101 tgaacattagattttaaaggttcaaaaggttcaaaacccattatctctcttct 3160
Db 35950 TGAACATTAGATATTTAAAGCTCAAGAGGTTCAAAATCCAACTCATATCTCTCTTCT 36009
QY 3161 ttcaactccctgctctctccctatatattactgattg-actgaacagatggtcccaa- 3218
Db 36010 TTCACCT-CCCTGCTCTCTCCCTATATTAATTACTGATTGCACTGAACAGCATGGTCCCAAT 36068
QY 3219 gatccagtcacaaatgagaacccagtcgctctctgtggtatcgtcgtcagactgctga 3278
Db 36069 GTAGCCATGCAAAATGAGAAACCCAGTGGCTCTTGTGTGTATGATGCAATGCAAGACTGCTGA 36128
QY 3279 agccag-agcatgactgattacccctcatggtggaagggaaccactcctggccttcgtg 3337
Db 36129 AGCAGAAGATGACTGATTACGCCTCATGGGTGAGGGGACCACCTCCTGGGCTTCGTG 36188
QY 3338 attgtcaggagcaagactgagatgctccctgccttcagtgctcctgcactcctcccttt 3397
Db 36189 ATTGTGAGGAGCAAGACTGAGATGCTCCCTGCCTTCAGTGTCTCTGCACTCTCCCTTT 36248
QY 3398 ctaatgaagatcaatagatttgcatttggaattggaattccaattaggaactcacatgttt 3457
Db 36249 CTAATGAAGATCCATAGATTTGCTACATTTGAGAAFTCCAAATAGGAACATCATGT 36308
QY 3458 tatctgcctatacaatttttaactgctgaaattaaagtatttttcaaaatcgtcctt 3517
Db 36309 TAUCTGCCCTATCAATTTTTTAACTGTCGTAATAATTAAGTTTTTCAAAATCTGCTCT 36368
QY 3518 gtaaatctctttcttacagtgctctgggatactatatacaacttgattctttgttaca 3577
Db 36369 GTAAATTAATTTTCTTACAGTGTCTGGCATACTATATCAACTTTGATTCCTTTGTACA 36428
QY 3578 acttt 3582
Db 36429 ACTTT 36433

RESULT 4
AL359314
LOCUS Homo sapiens chromosome 9 clone RP11-108L4, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.
ACCESSION AL359314
VERSION AL359314.12 GI:13396560
KEYWORDS HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173831)
Leongamornlert.D.
Direct Submission
Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:13277120.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA108L4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 173545 bases at least Q40
Consensus quality: 173599 bases at least Q30
Consensus quality: 173621 bases at least Q20
```

```
Insert size: 173631; sum-of-contigs
Insert size: 172123; 10.0% error; agarose-fp
Quality coverage: 10.35x in Q20 bases; sum-of-contigs Quality
coverage: 10.50x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 50595: contig of 50595 bp in length
* 50596 50695: gap of 100 bp
* 50696 157636: contig of 106941 bp in length
* 157637 157737: gap of 100 bp
* 157737 173831: contig of 16095 bp in length.
FEATURES
            Location/Qualifiers
            1..173831
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone_lib="RPC1-11.1"
            /clone_lib="RPC1-11.1"
            1..50595
            /note="assembly fragment:04137
            fragment_chain:1"
            50696..157636
            /note="assembly fragment:04236
            fragment_chain:1"
            157737..173831
            /note="assembly fragment:01656"
BASE COUNT 52107 a 36469 c 35018 g 50037 t 200 others
ORIGIN
```

```
Query Match      38.1%; Score 1363.4; DB 80; Length 173831;
Best Local Similarity 92.8%; Pred. No. 0;
Matches 1400; Conservative 0; Mismatches 106; Indels 3; Gaps 3;

QY 464  cagagatccctgggagaaatgccggccgccatcttgggtcatcgtagcctcgccctg 523
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156237 CACAGATCCTCGGAGAAATGCCGCCGCCATCTTGGGTCTCATGATGAGCTCGCCCTG 156296

QY 524  tcctgtgctccctgtgaggaagacattagaaaataaattgattgttctttaaagg 583
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156297 TGCTGTGCTCCCTTGTGAGGAAGGACATTAGAAATGAATGTGTCTCTTAAAGG 156356

QY 584  atgggcaggaacacagatcctgtgttggtatttattgaacgggattcacagatttgaa 643
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156357 ATGGGCAGGAAACACAGATCCTGTGTGTGGATATTTATTTCACGGGATTACAGATTTGAA 156416

QY 644  tgaagtcaaaagtgcagcattaccatgagaggaacacagagaaaaatcctgtatgct 703
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156417 TGAAGTCACAAAGTGCAGCATTTACCAATGAGGAGAAACACAGAGAAAAATCTTGATGGCT 156476

QY 704  tcacaagacatgcaacaaacaaatggaatactgtgatcacatgagcgagcagcaagctgg 763
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156477 TCACAGACATGCACAAACAAATGGAATACATGATGATGAGGAGCAAGCTGGG 156536

QY 764  gaggagataacccacgggagaggggtcaggattcttggcctgtcgtcctaaactgtgcgtt 823
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156537 GAGGAGATAACACACGGGCGAGAGGGTTCAGGATTTCTGCCCTCTGCTCTAACTGTGCGTT 156596

QY 824  cataaccaatacatttcatttcttaacccctcaaaacaaagctgtgtgtaatatcgtatct 883
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156597 CATAACCAAAATCATTTTCAATATTTTCTTAACCCCTCAAAACAAAGCTGTGTATATCTGATCT 156656

QY 884  ctacagttcctctggggcccaacttctccatataatccagccacacactcatttttaatt 943
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156657 CTACGGTTCCTTCTGGGGCCCAACATCTCTCATATATATCCAGCCACACTCATTTTAAATAT 156716
```

QY 944 tagttccacagatctgtactgtgacacctttctacactgtagaataaacattactactttgtt 1003
|||||
Db 156717 TAGTTCACAGATCTGTACTGTGACCTTTCTACACTGTAGAAATCAATTAATCTTTGTT 156776
QY 1004 caaagacccttctgttctgctcctaataatgttagctgactgttttttccctaagagtgctt 1063
|||||
Db 156777 CAAGACCCCTCGTGTGCTGCTTAATATGTAGTGAAGTGTGTTTCTTAAGGAGTGTCT 156836
QY 1064 gcccaggggactctgtgaacagctcgggaagcattcgaagatctcgaagatcttccaggttactt 1123
|||||
Db 156837 GCCCAGGGGATCTGTGACAGCGTGGGAAGCATCTCAAGATCTTTCAGAGGTTATCTT 156896
QY 1124 actagcacacagcatgatcattaccagagtggaattatctaatcaaacatcatctcagtgctc 1183
|||||
Db 156897 ACTAGCACACAGCATGATCATTTACGGAGTGAATTAATCAATCAATCATCTCTCAGTGTC 156956
QY 1184 ttgtcccatactgaattcatttcccaacttttgcacatttgcacatttccaaagacccaaatgtgc 1243
|||||
Db 156957 TTGTGCCATACAGAAATTCATTTCCCACTTTTGTGCCCAATTCGAAGACCTCAAAATGTC 157016
QY 1244 attccattaatatcacaggattaaacttttttttaacctggaagaattcaatgtttacat 1303
|||||
Db 157017 ATTCCATTAATATCAGAGATTAACCTTTTTTTTAACTGGAAGATTAATCAATGTATCAT 157076
QY 1304 gcagctatgggaatttaattacatatatttctgtttccagtgcaagatgactaagtccttt 1363
|||||
Db 157077 GCAGCTATGGGAATTTAATTACATATTTGTGTTTCCAGTGCAAGATGACTAAGTCTCTT 157136
QY 1364 atccctcccttctgttctgtattttttccagttataaagttaaaatgcttagctgtact 1423
|||||
Db 157137 ATTCCTCTCCCTTTGTGTGATTTTTCAGTATAAAGTTAAATGCTTTAGCTTTGTACT 157196
QY 1424 gaggctgtatacag-cacagcctctcccatccctccagccttctctgtcatcacatca 1482
|||||
Db 157197 GAGGCTGTATACAGCCAGAGCTCTCCCATCCCTCCAGCTTATCTGTCTATCACCATCA 157256
QY 1483 accctcccatcacactcaacaaactaaacttgtaattccttgaacatgtcaggacat 1542
|||||
Db 157257 ACCCTCTCCCAT-GACCTTAACAAATCTAAGTTGTAATTCCTTGAACATGTCAGG-CAT 157314
QY 1543 acattattctctgctgagagctctccttctgtctcttcttaacttagaataatgataag 1602
|||||
Db 157315 ACATATTCTCTTGTGCTGAGAGCTCTTCTTGTCTCTTAAATCTAGAAATGATGTAAG 157374
QY 1603 ttttgaataagtgtactatcttactctgcaagaagggacacatatagagattcatcat 1662
|||||
Db 157375 TTTTGAATTAAGTTGACTATCTTACTTCTATGCAAGAGGGACACATATGAGATTCTCAT 157434
QY 1663 cacatgagacagcaataactaaaagtgttaatttgattataaagagtttagataaataatg 1722
|||||
Db 157435 CACATGAGACAGCAAAATACTAAAAGTGAATTTGATTATAAGATTTAGATAAATATATG 157494
QY 1723 aaatgcaagagccacagaggaatgtttatggggacgcttttgaagcctggagtggaag 1782
|||||
Db 157495 AAATGCAAGAGCCACAGAGGGAAATGTTTATGGGCGACGTTTGTAAAGCCCTGGATGTAAG 157554
QY 1783 caaagcagggaaacctcatagtatcttataataataacttcttctctctctctctatc 1842
|||||
Db 157555 CAAGGACGGGAACCTCATAGTATCTTATATATATATCTTCTATCTCTCTATCTATCAC 157614
QY 1843 aatatcaacaagctttttcacagaaattcatgcaagtgcgaatacccccaagtaaacctttat 1902
|||||
Db 157615 AATATCCAACAAGCTTTTCACANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 157674
QY 1903 ccatttcatggtgagtcgcttagaattttggcaaatcactactggtcacttatctcaac 1962
Db 157675 NNN 157734
QY 1963 tttgagatg 1971
Db 157735 NNNNACAAG 157743

RESULT 5
AL358573/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
AL358573 267581 bp DNA HTG 15-APR-2001
Homo sapiens chromosome 9 clone RP11-133022, *** SEQUENCING IN
PROGRESS ***, 37 unordered pieces.
AL358573
AL358573.17 GI:13660951
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
McLay, K.
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13398774.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bal33022
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 256699 bases at least Q40
Consensus quality: 259744 bases at least Q30
Consensus quality: 261407 bases at least Q20
Insert size: 263881; sum-of-contigs
Insert size: 135491; 19.3% error; agarose-fp
Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality
coverage: 12.23x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 25718: contig of 25718 bp in length
* 25719 25818: gap of 100 bp
* 25819 29567: contig of 3749 bp in length
* 29568 29667: gap of 100 bp
* 29668 40520: contig of 10853 bp in length
* 40521 40620: gap of 100 bp
* 40621 50002: contig of 9382 bp in length
* 50003 50102: gap of 100 bp
* 50103 54942: contig of 4840 bp in length
* 54943 55042: gap of 100 bp
* 55043 59067: contig of 4025 bp in length
* 59068 59167: gap of 100 bp
* 59168 72327: contig of 13160 bp in length
* 72328 72427: gap of 100 bp
* 72428 79396: contig of 6969 bp in length
* 79397 79496: gap of 100 bp
* 79497 82738: contig of 3242 bp in length
* 82739 82838: gap of 100 bp
* 82839 87543: contig of 4705 bp in length
* 87544 87643: gap of 100 bp
* 87644 89847: contig of 2204 bp in length
* 89848 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7323 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length

* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2383 bp in length
* 121870 121969: gap of 100 bp
* 121970 125546: contig of 3577 bp in length
* 125547 125646: gap of 100 bp
* 125647 128990: contig of 3344 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length
* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2198 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.

FEATURES

source

1..267581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-133022"
/clone_lib="RPC1-11.1"
misc_feature 1..25718
/note="assembly_fragment:00645
fragment_chain:1
clone_end:T7
vector_side:left"
misc_feature 25819..29567
/note="assembly_fragment:01660
fragment_chain:1"
misc_feature 29668..40520
/note="assembly_fragment:03080
fragment_chain:1"
misc_feature 40621..50002
/note="assembly_fragment:00771
fragment_chain:1"
misc_feature 50103..54942
/note="assembly_fragment:02629
fragment_chain:1"
misc_feature 55043..59067
/note="assembly_fragment:04565
fragment_chain:1"

misc_feature 59168..72327
/note="assembly_fragment:00223
fragment_chain:1"
misc_feature 72428..79396
/note="assembly_fragment:03318
fragment_chain:1"
misc_feature 79497..82738
/note="assembly_fragment:00224
fragment_chain:2"
misc_feature 82839..87543
/note="assembly_fragment:03050
fragment_chain:2"
misc_feature 87644..89847
/note="assembly_fragment:01948
fragment_chain:3"
misc_feature 89948..92376
/note="assembly_fragment:02191
fragment_chain:3"
misc_feature 92477..99799
/note="assembly_fragment:00010"
misc_feature 99900..102117
/note="assembly_fragment:00672"
misc_feature 102218..109330
/note="assembly_fragment:01326"
misc_feature 109431..112187
/note="assembly_fragment:01498"
misc_feature 112288..116407
/note="assembly_fragment:01607"
misc_feature 116508..119386
/note="assembly_fragment:01899"
misc_feature 119487..121869
/note="assembly_fragment:02068"
misc_feature 121970..125546
/note="assembly_fragment:02469"
misc_feature 125647..128990
/note="assembly_fragment:02502"
misc_feature 129091..134055
/note="assembly_fragment:02669"
misc_feature 134156..138314
/note="assembly_fragment:02757"
misc_feature 138415..140612
/note="assembly_fragment:02776"
misc_feature 140713..143296
/note="assembly_fragment:03127"
misc_feature 143397..148113
/note="assembly_fragment:03634"
misc_feature 148214..151009
/note="assembly_fragment:03898"
misc_feature 151110..156616
/note="assembly_fragment:04220"
misc_feature 156717..158763
/note="assembly_fragment:04276"
misc_feature 158864..162129
/note="assembly_fragment:04418"
misc_feature 162230..164443
/note="assembly_fragment:04428"

Query Match 36.8%; Score 1319.8; DB 80; Length 267581;
Best Local Similarity 92.2%; Pred.No. 0;
Matches 1360; Conservative 0; Mismatches 112; Indels 3; Gaps 3;

QY 464 cagagatccctgggagaaatgccggccgcacattcttgggtcatcgatgagcttcgccctg 523
|| |||||
Db 165900 CACAGATCCCTGGGAGAAATGCCCGCCGCCCATCTTGGGTTCATCGATGAGCCTCGCCCTG 165841
QY 524 tgctgtgctccctgtgaggggaagacattagaaaaatgaattgattgttctcttaagg 583
|||||
Db 165840 TGCCTGTCCCGCTTGTGAGGGAAGGACATTAGAAAATGAATTGATGTGTTCTTTAAAGG 165781
QY 584 atgggcagaaacacatcctgtgtgatatattttaaaccgggattacagatttgaaa 643
|||||
Db 165780 ATGGCGCAGGAAACAGATCCTGTGTGGATATTTATTGAACGGGATTACAGATTTGAAA 165721

QY 644 tgaagtcaaaagtgcattaccatgagaggaagaaacagagcaagaaatcttgatggtc 703
|||||
Db 165720 TGAAGTCACAAAGTCGAGCATTTACCAATGAGAGGAAACAGAGCAAAATCTTGATGCT 165661
|||||
QY 704 tcaagacatgcacaacaaaaatggaactgtgagtgaatgagggcagccaagctggg 763
|||||
Db 165660 TCACAAAGACATGCAACAACAAAATGGAATACGTGTGATGACATGAGGAGCAAGCTGG 165601
|||||
QY 764 gaggagataaacacaggggcaagaggtcaggattctggccctgcctcaaacgtgcatt 823
|||||
Db 165600 GAGGAGATAAACCCGGGCGAGAGGTCTAGGATCTGGCCCTGCTGCCATAAATGTGCGTT 165541
|||||
QY 824 cataaaccatatttcattcttaacccctcaaaacaaagctgtgtaatatctgatct 883
|||||
Db 165540 CATAACCAAAATCATTTTCAATATTTCTTAACCCCTCAAAACAAAGCTGTGTAATATCTGATCT 165481
|||||
QY 884 ctacgggtctctctgggcccacatctctcatatataccagccacacactcatctttaaatt 943
|||||
Db 165480 CTACGGTTCCTTCTGGGCCCAACATCTCCATATATCCAGCCACACACTCATTTTAAATATT 165421
|||||
QY 944 tagttccagatctgtactgtgaacctttctacactgtagaataacattactatttgg 1003
|||||
Db 165420 TAGTTCCAGATCTGTACTGTGACCTTCTTACACGTGAGATTAACATTAATCTATTTGTT 165361
|||||
QY 1004 caaagacctctgtgtgctgctctaataatgtagctgactgttttctcctaagagtggtct 1063
|||||
Db 165360 CAAAGACCTTCGCTGCTGCTTAATATGTAGCTGACTGTGTTTCTTAAGGAGGTCTT 165301
|||||
QY 1064 ggcacggggatctgtgaacagctgggaagcatctcaagatctctccagggttatactt 1123
|||||
Db 165300 GGCCCAAGGGATCTGTGAACAGCTGGGAAGCATCTCAAGATCTTTCCAGGGTTATACTT 165241
|||||
QY 1124 actagcacacagcatgatcattacagagtggaattatctaatcaacatcatctcagtgctc 1183
|||||
Db 165240 ACTAGCACACAGCATGATCATTTACGGAGTGAATATCTTAATCAACATCATCTCTCAGTGTC 165181
|||||
QY 1184 ttgcccataactgaaattcatttcccactttgtgccattctcaagacctcaaaatgctc 1243
|||||
Db 165180 TTTGCCATACAGAAATTCATTTCCACATTTTGTGCCATTTCTCAAGACCTCAAAATGTC 165121
|||||
QY 1244 attccatttaataacagagattaaacttttttttaacctggaagaattcaatgttaact 1303
|||||
Db 165120 ATTCCATTAATATACAGAGATTAACTTTTATTTTAACTGGAAGAATTCATTTGTTACAT 165061
|||||
QY 1304 gcagctatgggaatttaattacattttgttttccagtgcaaaatgaactgaagctcttt 1363
|||||
Db 165060 GCAGCTATGGGAATTAATATACATTTTGTGTTTCCAGTGCAAAAGATGACTAAGTCCCTT 165001
|||||
QY 1364 atccctcccttgtgtgatttttttccagataaaagttaaagtcttagccttgact 1423
|||||
Db 165000 ATCCCTCCCTTTGTTGATTTTTCAGATATAAGTTAAATGCTTAGCCTTGACT 164941
|||||
QY 1424 gaggctgtatacag-cacagcctctcccatccctccagccttatctgtcatcacatca 1482
|||||
Db 164940 GAGGCTGTATACAGGCACAGCCTCTCCCATCCCTCCAGCCTTATCTGTCTATCACCATCA 164881
|||||
QY 1483 accctcccatcaccaactaaacaaatctaaacttaattcttcaaacatgtcagagact 1542
|||||
Db 164880 ACCCTCCCAT -GCATTAACAAATACTTAATCTGTAATCTTGAACATGTGAGG-CAT 164823
|||||
QY 1543 acattattctctgctgagagctcttctgtctctttaaacttagaatgataag 1602
|||||
Db 164822 ACATTAATCTCTGCTGAGAGCTCTTCTTGTCTTAACTAGATGATGATGAAG 164763
|||||
QY 1603 ttttgaataagttgactatcttacttcaatgcaaaagaggacacatatgagatcatcat 1662
|||||
Db 164762 TTTTGAATAAGTTGACTATCTTACTTCAATGCAAAAGAGGACACATATCAGATTTCATCAT 164703
|||||
QY 1663 cacatgagacagcaataactaaagtgtaatttgattataagagtttagataaataatg 1722
|||||
Db 164702 CACATGAGACAGCAATACTAAAGTGTAATTTGATTATTAAGAGTTTGTAGATAAATATAG 164643
|||||
QY 1723 aaatgcaagagccacagaggggaatgtttatggggcacgttttgaagcctgggagtggaag 1782

|||||
Db 164642 AAATGCAAGAGCCACAGAGGAATGTTTATGGGCGACCTTTGTAGCCCTGGGATGTGAAG 164583
|||||
QY 1783 caaagcagggaacctatagatattatataataataacttcttctctatctctacac 1842
|||||
Db 164582 CAAAGCGAGGAAACCTCATAGTATCTTATATAATATATTTNNNNNNNNNNNNNNNN 164523
|||||
QY 1843 aatatccacaagcttttcacagaattcatgcagtgcaaatccccaaaggtaaacctttat 1902
|||||
Db 164522 NNN 164463
|||||
QY 1903 ccatttcagtgtagcgctttagaattggca 1937
|||||
Db 164462 NNN 164428
|||||

RESULT 6

G56925/c G56925 546 bp DNA STS 30-MAR-2000
LOCUS SHGC-102485 Human Homo sapiens STS genomic, sequence tagged site.
DEFINITION G56925
ACCESSION G56925
VERSION G56925.1 GI:6122094
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 546)
AUTHORS Olivier, M. and Cox, D.R.
TITLE Unpublished, Olivier, M., Cox, D.R. (2000)
JOURNAL Unpublished (2000)
COMMENT

Contact: Michael Olivier, David R. Cox
Stanford Human Genome Center
Stanford University School of Medicine
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
Tel: (650) 320-5800
Fax: (650) 320-5801
Email: olivier@hgsc.stanford.edu
Primer A: TTAAGGATCTCAGACATGCGCTG
Primer B: TCATTACCAGGTTGGAGAGGA
STS size: 300
PCR Profile:
Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplifrag Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul

Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES source

Location/Qualifiers
1..546
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="9"
/clone_lib="Human"
29..328
primer_bind
29..51
complement(306..328)


```
QY 2783 aacgatgtctgagatccttaataaagaagaaccagtgatcgtcatgagtggaattc 2835
|||||
Db 62 AACGATGCTGAGATCCTTAATAAAGGAAACCAGTGTCATGAGTTGAATTC 10

RESULT 8
AX018075 359 bp DNA PAT 07-SEP-2000
LOCUS AX018075
DEFINITION Sequence 23 from Patent WO9946374.
ACCESSION AX018075
VERSION AX018075.1 GI:10042526
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 359)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
TITLE Human nucleic acid sequences from prostate tumour tissue
JOURNAL Patent: WO 9946374-A 23 16-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..359
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 121 a 75 c 94 g 69 t
ORIGIN

Query Match 6.6%; Score 235.4; DB 9; Length 359;
Best Local Similarity 99.6%; Pred. No. 6.7e-45;
Matches 236; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 281 cagggtgagaataagaagctgctgacttaccatctgagggccacacatctgctgaa 340
|||||
Db 123 CAGAGGTGAGAAATAAGAAGCTGCTGACTTTACCATCTGAGGCCACACATCTGCTGAA 182

QY 341 atggagataataacatcactcagaacagcaagatgacaataataatgcttaagtgtgac 400
|||||
Db 183 ATGGAGATAATTAACATCACTACTAGAACACAGATGACAAATAATAATGCTTAAGTAGTGAC 242

QY 401 atgttttcacattccagccctttaataatccacacacagagaagcacaaaaggaa 460
|||||
Db 243 ATGTTTTTCACATTTCAGCCCTTTAATAATATCCACACACAGGAGCAACAAAGGAA 302

QY 461 gcacagagatccctgggagaataagccggccgccatcttgggtcatcgtatgagcctc 517
|||||
Db 303 GCACAGAGATCCCTGGGAGAAATAGCCCGCCGCTTTGGGTTCATCGATGAGCCTC 359

RESULT 9
AL161625 143675 bp DNA HTG 20-JAN-2001
LOCUS AL161625
DEFINITION Homo sapiens chromosome 9 clone RP11-146P9, *** SEQUENCING IN
PROGRESS ***, 10 unordered pieces.
ACCESSION AL161625
VERSION AL161625.6 GI:9863607
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143675)
AUTHORS Plumb,B.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
COMMENT On Aug 21, 2000 this sequence version replaced gi:8894260.
----- Genome Center
```

```
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA146P9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 138647 bases at least Q40
Consensus quality: 140619 bases at least Q30
Consensus quality: 141706 bases at least Q20
Insert size: 142775; sum-of-contigs
Insert size: 147523; 5.9% error; agarose-fp
Quality coverage: 4.29x in Q20 bases; sum-of-contigs Quality
coverage: 4.22x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 12163: contig of 12163 bp in length
* 12164 12263: gap of 100 bp
* 12264 25269: contig of 13006 bp in length
* 25270 25369: gap of 100 bp
* 25370 57709: contig of 32340 bp in length
* 57710 57809: gap of 100 bp
* 57810 66792: contig of 8983 bp in length
* 66793 66892: gap of 100 bp
* 66893 102772: contig of 35880 bp in length
* 102773 102872: gap of 100 bp
* 102873 106863: contig of 3991 bp in length
* 106864 106963: gap of 100 bp
* 106964 129804: contig of 22841 bp in length
* 129805 129904: gap of 100 bp
* 129905 132223: contig of 2319 bp in length
* 132224 132323: gap of 100 bp
* 132324 135764: contig of 3441 bp in length
* 135765 135864: gap of 100 bp
* 135865 143675: contig of 7811 bp in length.
* Location/Qualifiers
1..143675
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-146P9"
/clone_lib="RPC1-11.1"
1..12163
/note="assembly_fragment:00440
clone_end:T7
vector_side:left"
12264..25269
/note="assembly_fragment:00434
fragment_chain:1"
25370..57709
/note="assembly_fragment:00526
fragment_chain:1"
57810..66792
/note="assembly_fragment:00661
fragment_chain:1"
66893..102772
/note="assembly_fragment:00323"
102873..106863
/note="assembly_fragment:00860"
106964..129804
/note="assembly_fragment:00978"
129905..132223
/note="assembly_fragment:01068"
```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

NOTE: This is a 'working draft' sequence. It currently consists of 37 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence. As soon as it is available and the accession number will be preserved.

FEATURES

source

Location/Qualifiers

1..267581

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/clone="RP11-133022"

/clone_lib="RPC1-11.1"

1..25718

/note="assembly_fragment:00645"

fragment_chain:1

clone_end:T7

vector_side:left

25819..29567

/note="assembly_fragment:01660"

fragment_chain:1

29668..40520

/note="assembly_fragment:03080"

fragment_chain:1

40621..50002

/note="assembly_fragment:00771"

fragment_chain:1

50103..54942

/note="assembly_fragment:02629"

fragment_chain:1

55043..59067

/note="assembly_fragment:04565"

fragment_chain:1

59168..72327

/note="assembly_fragment:00223"

fragment_chain:1

72428..79396

/note="assembly_fragment:03318"

fragment_chain:1

79497..82738

/note="assembly_fragment:00224"

fragment_chain:2

82839..87543

/note="assembly_fragment:03050"

fragment_chain:2

87644..89847

/note="assembly_fragment:01948"

fragment_chain:3

89948..92376

/note="assembly_fragment:02191"

fragment_chain:3

92477..99799

/note="assembly_fragment:00010"

99900..102117

/note="assembly_fragment:00672"

102218..109330

/note="assembly_fragment:01326"

109431..112187

/note="assembly_fragment:01498"

112288..116407

/note="assembly_fragment:01607"

116508..119386

/note="assembly_fragment:01899"

119487..121869

/note="assembly_fragment:02068"

121970..125546

/note="assembly_fragment:02469"

125647..128990

/note="assembly_fragment:02502"

129091..134055

/note="assembly_fragment:02669"

134156..138314

/note="assembly_fragment:02757"

138415..140612

/note="assembly_fragment:02776"

140713..143296

/note="assembly_fragment:03127"

143397..148113

/note="assembly_fragment:03634"

misc_feature

148214..151009

/note="assembly_fragment:03898"

151110..156616

/note="assembly_fragment:04220"

156717..158763

/note="assembly_fragment:04276"

158864..162129

/note="assembly_fragment:04418"

162230..164443

/note="assembly_fragment:04428"

Query Match

3.4%; Score 121; DB 80; Length 267581;

Best Local Similarity 100.0%; Pred. No. 7.2e-18;

Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acagaagaatagcaagtgcgcagagctgcacagaaaacagagggagatttgt 60

|||||

Db 253253 ACAGAGAATAAGCAAGTCCGAGAGCTGGCATCAGAAAACAGAGGGAGATTGTGT 253312

|||||

QY 61 ggctgcagccgagggagaccagagatctgcattggtggagggacctgatatacagag 120

|||||

Db 253313 GCCTGCAGCCGAGGAGACCAGAGATCTGCATGCTGGAGGACCTCATCATACAGAG 253372

|||||

QY 121 g 121

Db 253373 G 253373

RESULT 11

AF279290 580 bp DNA PRI 04-DEC-2000

LOCUS Homo sapiens prostate-cancer-specific DD3 protein gene, promoter

DEFINITION region and exon 1.

ACCESSION AF279290

VERSION AF279290.1 GI:11528086

KEYWORDS human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 580)

AUTHORS Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and Bussemakers,M.J.G.

TITLE Isolation and Characterization of the Promoter of the Human Prostate Cancer-specific DD3 Gene

JOURNAL J. Biol. Chem. 275 (48), 37496-37503 (2000)

PUBMED 10982808

REFERENCE 2 (bases 1 to 580)

AUTHORS Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and Bussemakers,M.J.G.

TITLE Direct Submission

JOURNAL Submitted (16-JUN-2000) Urology Research Laboratory, University Medical Center Nijmegen, Geert Grooteplein Zuid 10, Nijmegen 6525 GA, The Netherlands

FEATURES

Location/Qualifiers

1..580

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/map="9q21-q22"

1..460

260..288

/bound_moiety="high mobility group I Y protein"

461..5580

/product="prostate-cancer-specific DD3 protein"

461..580

/number=1

BASE COUNT 190 a 97 c 153 g 140 t

ORIGIN

Query Match

3.4%; Score 120; DB 89; Length 580;

Best Local Similarity 100.0%; Pred. No. 1.1e-17;

Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acagagaataagcaagtccgagaaagctgcacagaaaaacagagggagattgtgt 60
 |||||
 Db 461 ACAGAGAATAAGCAAGTCCGAGAAAGCTGCATCAGAAAAACAGAGGGAGATTGTGT 520
 |||||

QY 61 ggctgcagccgagggagaccagagaagatctgcattggtgggaagacacctgatgacagag 120
 |||||
 Db 521 GGCTGCAGCCGAGGAGACACAGGAAGATCTGCATGTTGGGAAGGACCTGATGATACAGAG 580
 |||||

RESULT 12

AC009556/c

LOCUS

DEFINITION Homo sapiens clone RP11-57C21, WORKING DRAFT SEQUENCE, 26 unordered
 pieces.

AC009556

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 172298)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,

Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., DeArillano, K., Depayre, E., Devon, K., Dewar, K.,

Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,

Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,

Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,

Karatas, A., Lenocksky, J., Lieu, C., Locke, K., MacDonald, P.,

Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,

Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,

Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,

Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,

Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,

Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

Direct Submission

Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 28, 2000 this sequence version replaced gi:6479158.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1672

Center clone name: 57C.21

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator; 5% of reads

Assembly: Dye-terminator Big Dye; 95% of reads

Consensus quality: 136385 bases at least Q40

Consensus quality: 156543 bases at least Q30

Consensus quality: 165534 bases at least Q20

Insert size: 165000; agarose-gel

Insert size: 169798; sum-of-contigs

Quality coverage: 4.2 in Q20 bases; agarose-gel

Quality coverage: 4.1 in Q20 bases.

NOTE: This is a 'working draft' sequence. It currently

consists of 26 contigs. The true order of the pieces

is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1016: contig of 1016 bp in length
 1017 1116: gap of 100 bp
 1117 2136: contig of 1020 bp in length
 2137 2236: gap of 100 bp
 2237 3571: contig of 1335 bp in length
 3572 3671: gap of 100 bp
 3672 5272: contig of 1601 bp in length
 5273 5372: gap of 100 bp
 5373 6381: contig of 1009 bp in length
 6382 6481: gap of 100 bp
 6482 8508: contig of 2027 bp in length
 8509 8608: gap of 100 bp
 8609 9886: contig of 1278 bp in length
 9887 9986: gap of 100 bp
 9987 11943: contig of 1957 bp in length
 11944 12043: gap of 100 bp
 12044 15112: contig of 3069 bp in length
 15113 15212: gap of 100 bp
 15213 19323: contig of 4111 bp in length
 19324 19423: gap of 100 bp
 19424 24065: contig of 4642 bp in length
 24066 24165: gap of 100 bp
 24166 31658: contig of 7493 bp in length
 31659 31758: gap of 100 bp
 31759 38737: contig of 6979 bp in length
 38738 38837: gap of 100 bp
 38838 46131: contig of 7294 bp in length
 46132 46231: gap of 100 bp
 46232 52344: contig of 6113 bp in length
 52345 52444: gap of 100 bp
 52445 59763: contig of 7319 bp in length
 59764 59863: gap of 100 bp
 59864 66816: contig of 6953 bp in length
 66817 66916: gap of 100 bp
 66917 74010: contig of 7094 bp in length
 74011 74110: gap of 100 bp
 74111 80599: contig of 6489 bp in length
 80600 80699: gap of 100 bp
 80700 89507: contig of 8808 bp in length
 89508 89607: gap of 100 bp
 89608 100337: contig of 10730 bp in length
 100338 100437: gap of 100 bp
 100438 114298: contig of 13861 bp in length
 114299 114398: gap of 100 bp
 114399 128530: contig of 14132 bp in length
 128531 128630: gap of 100 bp
 128631 142917: contig of 14287 bp in length
 142918 143017: gap of 100 bp
 143018 156243: contig of 13226 bp in length
 156244 156343: gap of 100 bp
 156344 172298: contig of 15955 bp in length.

FEATURES

Source

1. 172298
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-57C21"
 /clone_lib="RPCI-11 Human Male BAC"
 misc_feature
 1. 1016
 /note="assembly_fragment"
 misc_feature
 1117. 2136
 /note="assembly_fragment"
 misc_feature
 2237. 3571
 /note="assembly_fragment"
 misc_feature
 3672. 5272
 /note="assembly_fragment"
 misc_feature
 5373. 6381
 /note="assembly_fragment"
 misc_feature
 6482. 8508

misc_feature /note="assembly_fragment"
8609. .9886
/note="assembly_fragment"
9987. .11943
/note="assembly_fragment"
12044. .15112
/note="assembly_fragment"
15213. .19323
/note="assembly_fragment"
19424. .24065
/note="assembly_fragment"
24166. .31658
/note="assembly_fragment"
31759. .38737
/note="assembly_fragment"
38838. .46131
/note="assembly_fragment"
46232. .52344
/note="assembly_fragment"
52445. .59763
/note="assembly_fragment"
59864. .66816
/note="assembly_fragment"
66917. .74010
/note="assembly_fragment"
clone_end:T7
vector_side:left
74111. .80599
/note="assembly_fragment"
clone_end:SP6
vector_side:left
80700. .89507
/note="assembly_fragment"
89608. .100337
/note="assembly_fragment"
100438. .114298
/note="assembly_fragment"
114399. .128530
/note="assembly_fragment"
128631. .142917
/note="assembly_fragment"
143018. .156243
/note="assembly_fragment"
156344. .172298
/note="assembly_fragment"
BASE COUNT 51330 a 34914 c 34487 g 49065 t 2502 others
ORIGIN

Query Match 3.3%; Score 119.4; DB 61; Length 172298;
Best Local Similarity 99.2%; Pred. No. 1.7e-17;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 acagaagaatagcaagtccagagctgcatcagacagagagagagatttgtgt 60
Db 16561 ACAGAAGAATAGCAAGTCCGAGAGCTGGCATCAGAAAAACAGAGGGGAGATTGTGT 16502
QY 61 gctcagccagagagaccagagatctgcatggtggaaggacctgatatcaga 120
Db 16501 GCTCAGCCGAGGAGACCGAGAGATTTCATGGTGGGAAGACCTGATGATACAGAG 16442
QY 121 g 121
Db 16441 G 16441
RESULT 13
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
1. .7218
source
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
1. .7218
source
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match 2.2%; Score 77.4; DB 10; Length 7218;
Best Local Similarity 2.3%; Pred. No. 1.3e-07;
Matches 9; Conservative 245; Mismatches 131; Indels 0; Gaps 0;
QY 2830 gaattctctattatggatgcttagcttgcctcctctgctcctcttgacacatat 2889
Db 1064 GATYVV 1123
QY 2890 tagcttctagcctttgcttccagcactttatctttctccacacatcgcttacaatc 2949
Db 1124 YVV 1183
QY 2950 ctctctgctctgtgttggacttcccacacagaattccaacgactctcaagtcttt 3009
Db 1184 YVV 1243
QY 3010 tttccatcccccacacactgaatgcttagcctagacccttattttatattccaata 3069
Db 1244 YVV 1303
QY 3070 gatgctgctatgggctaattgctttagatgaacattagatatattaaagctaaagg 3129
Db 1304 YVV 1363
QY 3130 tcacaaatccacactattctctcttcttctccctccctgctcctcctcctatt 3189
Db 1364 YVV 1423
QY 3190 actgattgactgaacagagatggtcc 3214
Db 1424 YVVVVVVVVVGTACCAAAATCTTC 1448
RESULT 14
LOCUS I66494/c 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
1. .7218
source
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match 1.6%; Score 57.4; DB 10; Length 7218;
Best Local Similarity 4.3%; Pred. No. 0.0066;
Matches 16; Conservative 211; Mismatches 142; Indels 0; Gaps 0;
QY 444 aggaagcacaagaagcacagagatccctgggagaaatgccgcgccttcttgggt 503

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:53:08 ; Search time 46.06 Seconds
(without alignments)
146.495 Million cell updates/sec

Title: US-09-402-713a-7

Perfect score: 268

Sequence: 1 MFLHSSPKFKYPTQEAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:*

1: sp-archaea:*

2: sp-bacteria:*

3: sp-fungi:*

4: sp-human:*

5: sp-invertebrate:*

6: sp-mammal:*

7: sp-mhc:*

8: sp-organelle:*

9: sp-phage:*

10: sp-plant:*

11: sp-rodent:*

12: sp-unclassified:*

13: sp-vertebrate:*

14: sp-virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.5	22.6	1236	5 Q9NKF9	Q9NKF9 drosophila
2	60.5	22.6	1238	5 Q9VJW9	Q9VJW9 drosophila
3	60.5	22.6	1239	5 Q94902	Q94902 drosophila
4	60	22.4	187	2 Q9JUE3	Q9JUF3 neisseria m
5	60	22.4	218	2 Q9JRV5	Q9JRY5 neisseria m
6	59	22.0	1114	11 Q9JL73	Q9JL73 mus musculus
7	56.5	21.1	1061	2 P73998	P73998 synecocyst
8	56.5	21.1	6396	2 Q9KID7	Q9KID7 streptocyst
9	56	20.9	338	2 Q9RVU4	Q9RVU4 deinococcus
10	56	20.9	433	5 Q9V7N7	Q9V7N7 drosophila
11	56	20.9	598	5 Q9NFP1	Q9NFP1 drosophila
12	55.5	20.7	658	2 Q9FLJ5	Q9FLJ5 pseudomonas
13	55	20.5	168	14 Q9EL38	Q9EL38 bovine vira
14	55	20.5	196	5 Q9VKF5	Q9VKF5 drosophila
15	55	20.5	469	5 Q9GR75	Q9GR75 leishmania
16	55	20.5	1032	11 Q61989	Q61989 mus musculus
17	55	20.5	1935	5 Q9VQW0	Q9VQW0 drosophila
18	54.5	20.3	116	14 Q92886	Q92886 human immun
19	54.5	20.3	177	6 Q9XS95	Q9XS95 oryctolagus

20	54.5	20.3	244	2 Q9HYW1	Q9HYW1 pseudomonas
21	54.5	20.3	305	10 Q9M1C3	Q9M1C3 arabidopsis
22	54.5	20.3	461	11 Q9R1Y8	Q9R1Y8 mus musculus
23	54	20.1	204	3 Q9P3L8	Q9P3L8 neurospora
24	54	20.1	257	2 Q9Z8E3	Q9Z8E3 chlamydia p
25	54	20.1	450	10 Q9FMN6	Q9FMN6 arabidopsis
26	54	20.1	633	2 Q9HTU6	Q9HTU6 pseudomonas
27	54	20.1	784	4 Q00146	Q00146 homo sapien
28	53.5	20.0	250	8 Q31122	Q31122 schismocarp
29	53.5	20.0	348	8 Q35379	Q35379 paracyprih
30	53.5	20.0	408	4 Q9UGL4	Q9UGL4 homo sapien
31	53.5	20.0	421	8 Q9TIC2	Q9TIC2 balbiana i
32	53.5	20.0	443	8 Q32878	Q32878 palmaria pa
33	53.5	20.0	449	2 Q9Z9E0	Q9Z9E0 chlamydia p
34	53.5	20.0	783	4 Q9NXD0	Q9NXD0 homo sapien
35	53	19.8	671	5 Q9W365	Q9W365 drosophila
36	53	19.8	1161	11 Q9QYE7	Q9QYE7 rattus norv
37	53	19.8	2205	5 Q9NGQ2	Q9NGQ2 dictyosteli
38	52.5	19.6	216	2 Q9RVK0	Q9RVK0 deinococcus
39	52.5	19.6	357	2 Q9I720	Q9I720 pseudomonas
40	52.5	19.6	360	4 Q9H9N9	Q9H9N9 homo sapien
41	52.5	19.6	384	8 Q32401	Q32401 halosaccen
42	52.5	19.6	392	8 Q31855	Q31855 audouinella
43	52.5	19.6	424	8 Q9MV15	Q9MV15 ballia call
44	52.5	19.6	488	8 Q9THF8	Q9THF8 palmaria pa
45	52.5	19.6	493	4 Q9UF35	Q9UF35 homo sapien

ALIGNMENTS

RESULT 1

Q9NKF9 PRELIMINARY; PRT: 1236 AA.

AC Q9NKF9; 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)

DE KUZ PROTEIN.

GN KUZ OR BG:DS07660.3 OR CG7147.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Y, CN BW SP;

RX MEDLINE=99403001; PubMed=10471707;

RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T., Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D., Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A., Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K., Ceiniker S., Rubin G.M.;

RA "An exploration of the sequence of a 2.9-Mb region of the genome of Drosophila melanogaster: the Adh region.";

RL Genetics 153:179-219(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Y, CN BW SP;

RA Ceiniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G., Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M., Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A., Lee B., Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B., Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M., Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A., Sethi H., Snir E., Svirska R.R., Wan K.H., Weinburg T., Zhang R., Zieran L.L., Rubin G.M.;

RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE003407; AAF44800.1;

DR Flybase; FBgn0015954; kuz.

DR InterPro; IPR000130;

DR InterPro; IPR001590;

```

DR InterPro; IPR001762; -.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50214; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR SMART; SM00050; DISIN; 1.
SQ SEQUENCE 1236 AA; 136343 MW; 40461AC17040C9AD CRC64;

Query Match          22.6%; Score 60.5; DB 5; Length 1236;
Best Local Similarity 47.2%; Pred. No. 19;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPHTQEAQKQASLQEM----PGRHLGSSMSLA 40
Db 338 FKYPH-QKYTKANFAEGAFYDPSTGRRLGSSANVA 372
||||| I: ||| : | |||| :||

RESULT 2
Q9VJW9 PRELIMINARY; PRT; 1238 AA.
AC Q9VJW9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE KUZ PROTEIN.
GN KUZ OR BG:DS07660.3 OR CG7147.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Abmayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.K., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).

```

```

DR EMBL; AE003640; AAF53318.1; -.
DR HSP; P17494; 1KST.
DR FlyBase; FBgn0015954; kuz.
DR InterPro; IPR000130; -.
DR InterPro; IPR001590; -.
DR InterPro; IPR001762; -.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50214; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR SMART; SM00050; DISIN; 1.
SQ SEQUENCE 1238 AA; 136407 MW; 96D924B6188AC472 CRC64;

Query Match          22.6%; Score 60.5; DB 5; Length 1238;
Best Local Similarity 47.2%; Pred. No. 19;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPHTQEAQKQASLQEM----PGRHLGSSMSLA 40
Db 340 FKYPH-QKYTKANFAEGAFYDPSTGRRLGSSANVA 374
||||| I: ||| : | |||| :||

RESULT 3
Q94902 PRELIMINARY; PRT; 1239 AA.
AC Q94902;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE KUZBANIAN.
GN KUZ OR BG:DS07660.3 OR CG7147.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96355647; PubMed=8703057;
RA Rooke J., Pan D., Xu T., Rubin G.M.;
RT "KUZ, a conserved metalloprotease-disintegrin protein with two roles
RT in Drosophila neurogenesis.";
RL Science 273:1227-1231(1996).
DR EMBL; U60591; AAC47275.1; -.
DR HSP; P17494; 1KST.
DR MEROPS; M12.211; -.
DR FlyBase; FBgn0015954; kuz.
DR InterPro; IPR000130; -.
DR InterPro; IPR001590; -.
DR InterPro; IPR001762; -.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50214; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR SMART; SM00050; DISIN; 1.
KW Neurogenesis.
SQ SEQUENCE 1239 AA; 136522 MW; 7FCB514A1352577A CRC64;

Query Match          22.6%; Score 60.5; DB 5; Length 1239;
Best Local Similarity 47.2%; Pred. No. 19;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPHTQEAQKQASLQEM----PGRHLGSSMSLA 40
Db 341 FKYPH-QKYTKANFAEGAFYDPSTGRRLGSSANVA 375
||||| I: ||| : | |||| :||

RESULT 4
Q9JUF3 PRELIMINARY; PRT; 187 AA.
ID Q9JUF3
AC Q9JUF3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

```


DR	Pfam; PF00873; ACR_tran; 1.
DR	PRINTS; PR00702; ACRIFLAVINRP.
DR	PROSITE; PS00018; EF_HAND; UNKNOWN_1.
SQ	SEQUENCE 1061 AA; 114935 MW; 480D9B60FC09F8E2 CRC64;
Query Match 21.1%; Score 56.5; DB 2; Length 1061;	
Best Local Similarity 31.0%; Pred.No. 55;	
Matches	13; Conservative 11; Mismatches 17; Indels 1; Gaps
QY	5 ISSPFKYPHTQEAKQRSLGEMPGRHLSMSLALCLVPL 46 : : : : : : : : :
Dd	418 ISRFIKVKH-EDPOEAQAQAMGLTGAVIATSLVMVAFTPV 458 : : : : : :
RESULT 8	
Q9KID7	PRELIMINARY; PRT; 6396 AA.
ID	Q9KID7
AC	Q9KID7;
DT	01-OCT-2000 (Tremblrel. 15, Created)
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE	FREA.
DE	FEBA.
GN	FREA.
OS	Streptomyces hygroscopicus var. ascomyceticus.
OC	Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae;
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX	NCB1_TaxID=132248;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=20323220; PubMed=10863099;
RA	Ku K., Chung L., Revell W.P., Katz L., Reeves C.D.;
RT	"The FK520 gene cluster of streptomycetes hygrosopicus var.
RT	ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual
RT	polyketide extender units.";
RL	Gene 251:81-90(2000).
DR	EMBL; AF235504; AAF86396.1; -.
DR	InterPro; IPR000255; -.
DR	InterPro; IPR000794; -.
DR	InterPro; IPR001066; -.
DR	InterPro; IPR001227; -.
DR	InterPro; IPR001899; -.
DR	InterPro; IPR002085; -.
DR	Pfam; PF00107; adh_zinc; 2.
DR	Pfam; PF00109; ketoacyl-synt; 4.
DR	Pfam; PF00550; pp-binding; 4.
DR	Pfam; PF00698; Acyl-transf; 4.
DR	PROSITE; PS00075; ACP_DOMAIN; 4.
DR	PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_3.
DR	PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
DR	PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW	Phosphopantetheine; transferase.
SQ	SEQUENCE 6396 AA; 669881 MW; FF13BC2431D2E805 CRC64;
Query Match 21.1%; Score 56.5; DB 2; Length 6396;	
Best Local Similarity 35.6%; Pred.No. 3.9e+02;	
Matches	16; Conservative 6; Mismatches 14; Indels 9; Gaps
QY	12 PHT-----EQAEKAQRSIG---EMPGRHLGSSMSLALCLVPLV 47 : : : :
Dd	653 PHSVLGSDEEVLEAARQLGIHHRLPTRHAGHSERMQPLVPALL 697 : : : :
RESULT 9	
Q9RVU4	PRELIMINARY; PRT; 338 AA.
ID	Q9RVU4
AC	Q9RVU4;
DT	01-MAY-2000 (Tremblrel. 13, Created)
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT	01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE	ENDONUCLEASE III, PUTATIVE.

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusgkern D.R., Pacled J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RT Science 287:2185-2195(2000).
DR EMBL: AE003807; AAF58013.1; --
DR FlyBase: FBgn0020611; Vha44.
SQ SEQUENCE 433 AA; 46620 MW; FD29F44E6F317C46 CRC64;

Query Match 20.9%; Score 56; DB 5; Length 433;
Best Local Similarity 40.7%; Pred. No. 24;
Matches 11; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 4 HISSPFKYPHTQEAQKAQRSLGEMPG 30
DB 354 HHSHPHQHPHTNPLQSPVQKNSDEG 380

RESULT 11
Q9NFF1 ID Q9NFF1 PRELIMINARY; PRT; 558 AA.
AC Q9NFF1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE VANIN-LIKE PROTEIN.
GN VANIN-LIKE OR CG3648.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Galland F.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432008; PubMed=10501839;
RA Granjeaud S., Naquet P., Galland F.;
RT "An ESP description of the new Vanin gene family conserved from fly to human.";
RT Immunogenetics 49:964-972(1999).
DR EMBL: AJ276261; CAB77020.1; --
DR FlyBase: FBgn0040069; vanin-like.
DR InterPro: IPR003010; --
SQ SEQUENCE 558 AA; 62341 MW; 7920621923BB7779 CRC64;

Query Match 20.9%; Score 56; DB 5; Length 558;
Best Local Similarity 33.3%; Pred. No. 32;
Matches 12; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 3 LHISPFKYPHTQEAQKAQRSLGEMPGRHGSSMS 38
DB 292 IYVAQVPKPYTRSLKRAKRSLOBIQTRQVASSSS 327

RESULT 12
Q9FLJ5 ID Q9FLJ5 PRELIMINARY; PRT; 698 AA.
AC Q9FLJ5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PUTATIVE HRPO.
GN PUTATIVE HRPO.
OS Pseudomonas glumae.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=337;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM4;
RA Takikawa Y., Eo S., Adachi S., Kojima M.;
RT "Deduced hrpO genes from Burkholderia glumae and Acidovorax avenae subsp. avenae.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB053455; BAB20913.1; --
SQ SEQUENCE 698 AA; 75349 MW; 48B8E407FCE559A3 CRC64;

Query Match 20.7%; Score 55.5; DB 2; Length 698;
Best Local Similarity 27.8%; Pred. No. 48;
Matches 15; Conservative 10; Mismatches 20; Indels 9; Gaps 1;

QY 1 MFLHISPFKYPHTQEAQKAQRSLGEMPGRHGSS-----MSLALCLVP 45
DB 244 LLLSVAAGVMTITRVADERQAKQKRSLSGLGHLGSSRALFFAALLLAFAIVP 297

RESULT 13
Q9EI38 ID Q9EI38 PRELIMINARY; PRT; 168 AA.
AC Q9EI38;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Bovine viral diarrhoea virus-1.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Pestivirus.
OX NCBI_TaxID=121864;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26-V639;
RA Vilcek S., Paton D.J., Durkovic B., Strojny L., Ibata G., Moussa A.,
RA Voltsch A., Rossmanith W., Vega S., Scicluna M.T., Palfi V.;
RT "Bovine viral diarrhoea virus genotype 1 can be separated into at least eleven genetic groups.";
RT Arch. Virol. 0:0-0(2000).
DR EMBL: AF287281; AAG24564.1; --
FT NON_TER 168
SQ SEQUENCE 168 AA; 18947 MW; F69B9DE96BA2C66 CRC64;

Query Match 20.5%; Score 55; DB 14; Length 168;
Best Local Similarity 40.5%; Pred. No. 12;
Matches 15; Conservative 4; Mismatches 14; Indels 4; Gaps 2;

QY 3 LHISPFKYPHTQEAQKAQRSLGEMPGR---HLGSS 36
DB - 39 LHPOSTLKLPH-KRGEKVPKTSLSLPRKRGDCRLGNS 74

RESULT 14
Q9VKF5 ID Q9VKF5 PRELIMINARY; PRT; 196 AA.
AC Q9VKF5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM'protein - protein search, using sw model

Run on: July 30, 2001, 16:53:31 ; Search time 17.83 Seconds
(without alignments)
97.983 Million cell updates/sec

Title: US-09-402-713A-7

Perfect score: 268

Sequence: 1 MFLHISPPFKYPTQEAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	60.5	22.6	343	1	TA2R_BOVIN
2	59	22.0	1113	1	HDAS_MOUSE
3	59	22.0	1122	1	HDAS_HUMAN
4	55	20.5	1039	1	ITAA_MOUSE
5	54.5	20.3	461	1	COLA_BOVIN
6	54.5	20.3	461	1	COLA_HUMAN
7	54.5	20.3	461	1	COLA_MOUSE
8	53.5	20.0	446	1	GNTU_ECOLI
9	53.5	20.0	451	1	PTCC_BACST
10	53.5	20.0	580	1	5NTD_BOOMI
11	53	19.8	554	1	DCS4_GOSAR
12	52.5	19.6	441	1	RXRA_BRARE
13	52	19.4	337	1	RIR2_TRYBB
14	52	19.4	2774	1	MAFA_RAT
15	51.5	19.2	269	1	AQPL_HUMAN
16	51.5	19.2	470	1	RBL_PROHO
17	51.5	19.2	569	1	4CL2_ORISA
18	51	19.0	352	1	KE4_BRARE
19	51	19.0	1038	1	ITAA_HUMAN
20	51	19.0	1162	1	ITAD_HUMAN
21	51	19.0	2485	1	POLN_EEVV3
22	51	19.0	2492	1	POLN_EEVPV
23	50.5	18.8	409	1	AXIA_BRARE
24	50.5	18.8	554	1	DCS2_GOSAR
25	50.5	18.8	2472	1	NCR2_MOUSE
26	50	18.7	130	1	ITP_SCHGR
27	50	18.7	134	1	ITPL_SCHGR
28	50	18.7	456	1	HMFT_DROHY
29	50	18.7	467	1	RBL_PHORE
30	50	18.7	528	1	ANM3_RAT
31	50	18.7	1100	1	JAK3_RAT
32	50	18.7	1163	1	ITAX_HUMAN
33	50	18.7	2492	1	POLN_EEVTV

RESULT 1

TA2R_BOVIN					
ID	TA2R_BOVIN	STANDARD;	PRT;	343 AA.	
AC	Q95125; Q95124;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	THROMBOXANE A2 RECEPTOR (TXA2-R) (PROSTANOID TP RECEPTOR).				
GN	TBXA2R.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_TaxID=9913;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Liver, and Heart;				
RX	MEDLINE=98221872; PubMed=9561101;				
RA	Muck S., Schroer K.;				
RT	"Cloning, tissue-specific expression and regulation of the bovine				
RT	thromboxane A2 receptor";				
RL	Adv. Exp. Med. Biol. 433:47-50(1997).				
CC	-!- FUNCTION: RECEPTOR FOR THROMBOXANE A2 (TXA2), A POTENT STIMULATOR				
CC	OF PLATELET AGGREGATION. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED				
CC	BY A G-PROTEIN THAT ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND				
CC	MESENSENGER SYSTEM. IN THE KIDNEY, THE BINDING OF TXA2 TO GLOMERULAR				
CC	TP RECEPTORS CAUSES INTENSE VASOCONSTRICTION.				
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.				
CC	-!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; U53485; AAC34309.1; -				
DR	EMBL; U53484; AAC34308.1; -				
DR	GCRDB; GCR.1217; -				
DR	InterPro; IPR000276; -				
DR	InterPro; IPR001105; -				
DR	Pfam; PF00001; 7tm1; 1.				
DR	PRINTS; PR00429; THROMBOXANER.				
DR	PROSITE; PS00237; G-PROTEIN_RECEP_F1.1; 1.				
DR	PROSITE; PS00262; G-PROTEIN_RECEP_F1.2; 1.				
KW	G-protein coupled receptor; Transmembrane; Glycoprotein.				
FT	DOMAIN 1 29 EXTRACELLULAR (POTENTIAL).				
FT	TRANSMEM 30 52 1 (POTENTIAL).				
FT	DOMAIN 53 66 CYTOPLASMIC (POTENTIAL).				
FT	TRANSMEM 67 87 2 (POTENTIAL).				
FT	DOMAIN 88 106 EXTRACELLULAR (POTENTIAL).				
FT	TRANSMEM 107 128 3 (POTENTIAL).				
FT	DOMAIN 129 149 CYTOPLASMIC (POTENTIAL).				
FT	TRANSMEM 150 172 4 (POTENTIAL).				

P11510 rattus norv
P41558 methanococ
P93665 gossypium h
P41561 vibrio sp.
Q25637 periplaneta
O62816 ovis aries
P70597 rattus norv
P17169 e glucosami
P00892 escherichia
Q39761 gossypium a
Q921x9 helicobacte
Q14511 homo sapien

ALIGNMENTS

```

Query Match      22.0%; Score 59; DB 1; Length 1113;
Best Local Similarity 32.0%; Pred. No. 8.7;
Matches 16; Conservative 14; Mismatches 16; Indels 4; Gaps 2;

QY      4  HISSPFKYPTQEAQFAQRSL--GEMPRHGLGSSMSLALCLVPLVREG 50
          | | | | | | | | | | | | | | | | | | | | | | | |
DB      377  HLTAAPKSLSTQQAERQALQSLRQGGTLTKGFMST-SIPGCLLVGALEG 425

RESULT      3
HDA5_HUMAN  STANDARD;          PRT;      1122 AA.
ID      Q9UQL6; O60528; O60340;
DT      01-OCT-2000 (Rel. 40, Created)
DT      01-OCT-2000 (Rel. 40, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      HISTONE DEACETYLASE 5 (HD5) (ANTIGEN NY-CO-9).
DE      HDAC5 OR KIAA0600.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RP      SEQUENCE FROM N.A.
RP      MEDLINE=99238449; PubMed=10220385;
RX      Grozinger C.M., Hassig C.A., Schreiber S.L.;
RT      "Three proteins define a class of human histone deacetylases related
RT      to yeast Hda1p.";
RL      Proc. Natl. Acad. Sci. U.S.A. 96:4868-4873(1999).
RN      [2]
RP      SEQUENCE OF 189-1085 FROM N.A.
RP      TISSUE=Colorectal carcinoma;
RX      MEDLINE=98272252; PubMed=9610721;
RX      Scanlan M.J., Chen Y.-T., Williamson B., Gure A.O., Stockert E.,
RX      Gordon J.D., Tureci O., Sahin U., Pfreundschuh M., Old L.J.;
RT      "Characterization of human colon cancer antigens recognized by
RT      autologous antibodies.";
RL      Int. J. Cancer 76:652-658(1998).
RN      [3]
RP      SEQUENCE OF 407-1122 FROM N.A.
RP      TISSUE=Brain;
RX      MEDLINE=98290545; PubMed=9628581;
RX      Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RX      Nomura N., Ohara O.;
RT      "Prediction of the coding sequences of unidentified human genes. IX.
RT      The complete sequences of 100 new cDNA clones from brain which can
RT      code for large proteins in vitro.";
RL      DNA Res. 5:31-39(1998).
CC      -!- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
CC      THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
CC      HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
CC      REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS (BY
CC      SIMILARITY).
CC      -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC      -!- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC      FAMILY. HD SUBFAMILY 2.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF132608; RAD29047.1; -
DR      EMBL; AF039691; RAC18040.1; -
DR      EMBL; AB011172; BAA25526.1; -
DR      MIM; 605315; -
DR      InterPro; IPR000286; -
DR      Pfam; PF00850; Hist_deacetyl; 1.
DR      PRINTS; PR01270; HDASUPER.

```


GN COROIA OR CORO1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Spine;
RX MEDLINE=95278344; PubMed=7758584;
RA Suzuki K., Nishihata J., Arai Y., Honma N., Yamamoto K., Irimura T.,
RA Toyoshima S.;
RT "Molecular cloning of a novel actin-binding protein, p57, with a WD
RT repeat and a leucine zipper motif.";
RL FEBS Lett. 364:283-288(1995).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF
CC CHEMOTAXIS.
CC -1- SUBUNIT: BINDS ACTIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, THYMUS, SPLEEN, BONE
CC MARROW AND LYMPH NODE. LOW IN LUNG AND GUT.
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D44496; BAA07939.1; -
DR InterPro: IPR001680; -
DR Pfam; PF00400; WD40; 3.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 2.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Actin-binding; Repeat; WD repeat; Coiled coil.
FT REPEAT 79 119 WD 1.
FT REPEAT 129 169 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 218 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 424 460 COILED COIL (POTENTIAL).
FT SEQUENCE 461 AA; 50979 MW; 1721A5D093C1130A CRC64;

Query Match 20.3%; Score 54.5; DB 1; Length 461;
Best Local Similarity 43.8%; Pred. No. 13;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;
QY 2 FLHISPPFKYPHTQEAQKEAQRSLGEMPGRLH 33
DB 303 FLHYSMF-----SSKESQRCGMGYMPKRG 327

RESULT 6
COIA_HUMAN
ID COIA_HUMAN STANDARD; PRT; 461 AA.
AC P31146;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CORONIN-LIKE PROTEIN P57 (CORONIN 1A).
GN COROIA OR CORO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Periphereal blood leukocytes;
RX MEDLINE=95278344; PubMed=7758584;

RA Suzuki K., Nishihata J., Arai Y., Honma N., Yamamoto K., Irimura T.,
RA Toyoshima S.;
RT "Molecular cloning of a novel actin-binding protein, p57, with a WD
RT repeat and a leucine zipper motif.";
RL FEBS Lett. 364:283-288(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX Grodan A., Keep N.H., Reeves E., Segal A.W.;
RA Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Liau G., Popa I., Argraves K., Argraves W.S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 355-374.
RC TISSUE=Keratinocytes;
RX Rasmussen H.H., Van Damme J., Puype M., Gesser B., Celis J.E.,
RA Vandekerckhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
RT protein database of normal human epidermal keratinocytes.";
RL Electrophoresis 13:960-969(1992).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF
CC CHEMOTAXIS (BY SIMILARITY).
CC -1- SUBUNIT: BINDS ACTIN (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, THYMUS, SPLEEN, BONE
CC MARROW AND LYMPH NODE. LOW IN LUNG AND GUT.
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D44497; BAA07940.1; -
DR EMBL; X89109; CAA61482.1; -
DR EMBL; U34690; AAA7058.1; -
DR MIM; 605000; -
DR Aarhus/Ghent-2DPAGE; 2416; IEF.
DR InterPro: IPR001680; -
DR Pfam; PF00400; WD40; 3.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 2.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Actin-binding; Repeat; WD repeat; Coiled coil.
FT REPEAT 79 119 WD 1.
FT REPEAT 129 169 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 218 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 424 460 COILED COIL (POTENTIAL).
FT VARIANT 355 355 K -> D.
FT VARIANT /FTID-VAR_007722.
FT VARIANT S -> D.
FT VARIANT /FTID-VAR_007723.
FT CONFLICT 8 8 S -> T (IN REF. 3).
FT CONFLICT 245 245 R -> W (IN REF. 3).
SQ SEQUENCE 461 AA; 51026 MW; DE3FEDA57041515E CRC64;

Query Match 20.3%; Score 54.5; DB 1; Length 461;
Best Local Similarity 43.8%; Pred. No. 13;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;
QY 2 FLHISPPFKYPHTQEAQKEAQRSLGEMPGRLH 33
DB 303 FLHYSMF-----SSKESQRCGMGYMPKRG 327

```
RESULT 7
COLA_MOUSE
ID COLA_MOUSE STANDARD; PRT; 461 AA.
AC 089053; Q9R288;
DT 30-MAY-2000 (Rel. 39, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CORONIN-LIKE PROTEIN P57 (CORONIN 1A).
GN COROLA OR CORO1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98449467; PubMed-9778037;
RA Okumura M., Kung C., Wong S., Rodgers M., Thomas M.L.;
RT "Definition of family of coronin-related proteins conserved between
RT humans and mice: close genetic linkage between coronin-2 and CD45-
RT associated protein."
RL DNA Cell Biol. 17:779-787(1998).
RN [2]
RP SEQUENCE OF 156-276 FROM N.A.
RC STRAIN-BALB/C; TISSUE-Spleen;
RX MEDLINE-99012997; PubMed-9798653;
RA Chu C.C., Paul W.E.;
RT "Expressed genes in interleukin-4 treated B cells identified by cDNA
RT representational difference analysis."
RL Mol. Immunol. 35:487-502(1998).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF
CC CHEMOTAXIS (BY SIMILARITY).
CC -!- SURUNIT: BINDS ACTIN (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSION RESTRICTED TO HEMATOPOIETIC CELLS.
CC -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF143955; AAD32703.1; -
DR EMBL; U89399; AAC36506.1; -
DR MGD; MGI:1345961; Corola.
DR InterPro; IPR001680; -.
DR Pfam; PF00400; WD40. 1.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 2.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Actin-binding; Repeat; WD repeat; Coiled coil.
FT REPEAT 79 119 WD 1.
FT REPEAT 129 169 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 218 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 424 460 COILED COIL (POTENTIAL).
FT SEQUENCE 461 AA; 50975 MW; 51D48ED91E6FE82A CRC64;

RESULT 8
GNTU_ECOLI
ID GNTU_ECOLI STANDARD; PRT; 446 AA.
AC P46858; P76694; P76695;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE LOW-AFFINITY GLUCONATE TRANSPORTER (GLUCONATE PERMEASE) (GNT-I
DE SYSTEM).
GN GNTU OR B3435/B3436.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RX MEDLINE-97280784; PubMed-9135111;
RA Izu H., Adachi O., Yamada M.;
RT "Gene organization and transcriptional regulation of the gntRku
RT operon involved in gluconate uptake and catabolism of Escherichia
RT coli."
RL J. Mol. Biol. 267:778-793(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE-96236044; PubMed-8655507;
RA Tong S., Porco A., Isturiz T., Conway T.;
RT "Cloning and molecular genetic characterization of the Escherichia
RT coli gntR, gntK, and gntU genes of GntI, the main system for
RT gluconate metabolism."
RL J. Bacteriol. 178:3260-3269(1996).
CC -!- FUNCTION: PART OF THE GLUCONATE UTILIZATION SYSTEM GNT-I; LOW-
CC AFFINITY INTAKE OF GLUCONATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE GNTP FAMILY OF PERMEASES.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D84362; BAA12326.1; -
DR EMBL; U18997; AAA58234.1; ALT_FRAME.
DR EMBL; U18997; AAA58233.1; ALT_FRAME.
DR EMBL; AE000420; AAC76461.1; ALT_FRAME.
DR EMBL; AE000420; AAC76460.1; ALT_FRAME.
DR EcoGene; EGI2631; gntU.
KW Gluconate utilization; Sugar transport; Transport; Transmembrane;
KW Inner membrane; Multigene family.
FT TRANSMEM 2 22 POTENTIAL.
FT TRANSMEM 27 47 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 110 130 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.
```

```

Query Match          20.0%: Score 53.5; DB 1; Length 451;
Best Local Similarity 28.0%: Pred. No. 17;
Matches 14; Conservative 12; Mismatches 19; Indels 5; Gaps 1;

QY 5 ISSPEKYPHTQEAQKQAGSLGEMPRHLGS-----SMSLALCLVPLVRE 49
    ::||:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116 LATPYQVFTPEGAKETIMVSSGGIPVQWVSGKGLFVAMILAIVSTEIYRK 165
    ::||:| | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
5NTD_BOOMI          STANDARD;          PRT;          580 AA.
AC P52307; P90696;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE 5'-NUCLEOTIDASE PRECURSOR (EC 3.1.3.5) (FRAGMENT)..
OS Boophilus microplus (Cattle tick).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Boophilus.
ON NCBI_TaxID=6941;
RX [1]
SEQUENCE FROM N.A.
RX MEDLINE=99306777; PubMed=10380109;
RA Llyou N., Hamilton S., Elvin C., Willadsen P.;
RT "Cloning and expression of ecto 5-nucleotidase from the cattle tick
RL Boophilus microplus.";
RL Insect Mol. Biol. 8:257-266(1999).
RX [2]
SEQUENCE OF 15-40 AND 162-180.
RX MEDLINE=93250870; PubMed=8387372;
RA Willadsen P., Riding G.A., Jarney J., Atkins A.;
RT "The nucleotidase of Boophilus microplus and its relationship to
RL enzymes from the rat and Escherichia coli.";
RL Insect Biochem. Mol. Biol. 23:291-295(1993).
CC -1- FUNCTION: DEGRADATION OF EXTERNAL UDP-GLUCOSE TO URIDINE
CC MONOPHOSPHATE AND GLUCOSE-1-PHOSPHATE, WHICH CAN THEN BE USED BY
CC THE CELL.
CC -1- CATALYTIC ACTIVITY: UDP-SUGAR + H(2)O = UMP + SUGAR 1-PHOSPHATE.
CC -1- CATALYTIC ACTIVITY: A 5'-RIBONUCLEOTIDE + H(2)O = A
CC RIBONUCLEOSIDE + ORTHOPHOSPHATE.
CC -1- COFACTOR: ZINC.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- TISSUE SPECIFICITY: GUT, OVARIES AND SALIVARY GLANDS.
CC -1- PTM: GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U80634; AAB38963.1; -.
CC InterPro; IPR000934; -.
CC InterPro; IPR022224; -.
CC PROSITE; PS00785; 5_NUCLEOTIDASE.1; FALSE_NEG.
CC PROSITE; PS00786; 5_NUCLEOTIDASE.2; FALSE_NEG.
CC Pfam; PF01009; 5_nucleotidase; 1.
CC Hydrolase; Gpi-anchor; Glycoprotein; Signal; Zinc.
CC NON_TER 1 1
CC FT SIGNAL 1 14
CC FT CHAIN 15 552
CC FT PROPEP 553 580
CC FT LIPID 552 552
CC FT CARBOHYD 172 172
CC FT CARBOHYD 285 285
CC FT CARBOHYD 423 423
CC FT CARBOHYD 536 536
CC
CC 5'-NUCLEOTIDASE.
CC REMOVED IN MATURE FORM (BY SIMILARITY).
CC GPI-ANCHOR (POTENTIAL).
CC N-LINKED (GLCNAC. .) (PROBABLE).
CC N-LINKED (GLCNAC. .) (POTENTIAL).
CC N-LINKED (GLCNAC. .) (POTENTIAL).
CC N-LINKED (GLCNAC. .) (POTENTIAL).

```

```
FT CONFLICT 15 15 T -> K (IN REF. 2).
FT CONFLICT 37 39 SOT -> HXG (IN REF. 2).
SQ SEQUENCE 580 AA: 63460 MW: 588EEF2014071AB7 CRC64;

Query Match 20.0%; Score 53.5; DB 1; Length 580;
Best Local Similarity 36.1%; Pred. No. 23;
Matches 22; Conservative 5; Mismatches 21; Indels 13; Gaps 3;

QY 4 HISSPFKYPHTQEA--OKEAORSIGE-----MCRHLGSSMSLAIC-----LVPLVREG 50
Db 175 YLSSPGKVRFTDEACIOREAGRLRREGCVQVIAVGHSGVPRDLEICRVEPSLVVGG 234

QY 51 H 51
Db 235 H 235

RESULT 11
DCS4_GOSAR STANDARD; PRT; 554 AA.
AC Q49853;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE (+)-DELTA-CADINENE SYNTHASE ISOZYME C2 (EC 4.6.1.11) (D-CADINENE
DE SYNTHASE).
GN CADI-C2
OS Gossypium arboreum.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
OX NCBI_TaxID=29729;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NANKING;
RA Meng Y., Jia J., Liu C., Liang W., Zhou X., Heinstein P., Chen X.-Y.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: (E,E)-FARNESYL DIPHOSPHATE CYCLIZING.
CC -!- CATALYTIC ACTIVITY: 2-TRANS,6-TRANS-FARNESYL DIPHOSPHATE = (+)-
CC DELTA-CADINENE + DIPHOSPHATE.
CC -!- PATHWAY: FIRST COMMITTED STEP IN THE PATHWAYS LEADING TO THE
CC RELATED PHYTOALEXINS GOSSYPOL AND LACINILENE C.
CC -!- SIMILARITY: TO OTHER PLANT TERPENE CYCLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y16432; CAA76223.1; -
DR HSSP; Q40577; 5EAO.
DR InterPro; IPR001906; -
DR Pfam; PF01397; Terpene_synth; 1.
KW Lyase; Multigene family.
SQ SEQUENCE 554 AA: 64117 MW: 35DDDD66D3E838AAC CRC64;

Query Match 19.8%; Score 53; DB 1; Length 554;
Best Local Similarity 38.3%; Pred. No. 25;
Matches 18; Conservative 5; Mismatches 18; Indels 6; Gaps 2;

QY 4 HISSPFKYPHTQEAQKAQRAQSLGEMPGRLHSGMSLAICLVPLVREG 50
Db 490 HVESAWK-DVNGEFQKPT-----ENPTEVLNRLNRLARYMDVLYREG 530

RESULT 12
RXRA_BRARE STANDARD; PRT; 441 AA.
ID RXRA_BRARE
```

```
AC Q90416;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RETINOIC ACID RECEPTOR RXR-ALPHA.
GN RXRA OR NR2B1 OR RXR.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96009547; PubMed=7565671;
RA Jones B.B., Ohno C.K., Allenby G., Boffa M.B., Levin A.A.,
RA Grippo J.F., Petkovich M.;
RT "New retinoid X receptor subtypes in zebra fish (Danio rerio)
RT differentially modulate transcription and do not bind 9-cis retinoic
RT acid."
RL Mol. Cell. Biol. 15:5226-5234(1995).
CC -!- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS
CC 9-CIS RETINOIC ACID (9C-RA) (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29940; AAC59720.1; -
DR HSSP; P19793; 2NLL.
DR ZFIN; ZDB-GENE-980526-36; rxra.
DR InterPro; IPR000003; -
DR InterPro; IPR000536; -
DR InterPro; IPR001628; -
DR InterPro; IPR001723; -
DR Pfam; PF00104; hormone_rec; 1.
DR PRINTS; PR00105; zf-C4; 1
DR PRINTS; PR00047; STROIDFINGER.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00545; RETINOIDXR.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family.
FT DOMAIN 1 116 MODULATING (BY SIMILARITY).
FT DNA_BIND 117 182 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 117 137 C4-TYPE.
FT ZN_FING 153 177 C4-TYPE.
FT DOMAIN 183 206 HINGE.
FT DOMAIN 207 441 LIGAND-BINDING (BY SIMILARITY).
SQ SEQUENCE 441 AA: 48690 MW: E0F8F1EBE31CAEC CRC64;

Query Match 19.6%; Score 52.5; DB 1; Length 441;
Best Local Similarity 31.1%; Pred. No. 23;
Matches 14; Conservative 6; Mismatches 18; Indels 7; Gaps 1;

QY 2 FLHISPFKYPHTQEAQKAQRAQSLGEMPGRLHSGMSLS 39
Db 7 YLHLSLSLQVAHGLSLSPSPQPLGSMVSHHHHPHSLGLSPYSV 51

RESULT 13
RIR2_TRYBB STANDARD; PRT; 337 AA.
ID RIR2_TRYBB
```

AC O15910; O15880;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (EC 1.17.4.1)
DE (RIBONUCLEOTIDE REDUCTASE R2 SUBUNIT).
GN RNR2 OR NRDB.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OC Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=427;
RX MEDLINE=97338127; PubMed=9192674;
RA Hofer A., Schmidt P.P., Graslund A., Thelander L.;
RT "Cloning and characterization of the R1 and R2 subunits of
RT ribonucleotide reductase from Trypanosoma brucei";
RL Proc. Natl. Acad. Sci. U.S.A. 94:6959-6964(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97459778; PubMed=9315738;
RA Dornmeyer M., Schoneck R., Dittmar G.A.G., Krauth-Siegel R.L.;
RT "Cloning, sequencing and expression of ribonucleotide reductase R2
RT from Trypanosoma brucei";
RL FEBS Lett. 414:449-453(1997).
CC -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC -!- CATALYTIC ACTIVITY: 2'DEoxyRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC THIOREDOXIN + H(2)O -> RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC THIOREDOXIN.
CC -!- COFACTOR: CONTAINS TWO IRON IONS.
CC -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC SMALL CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U80911; AAB70705.1; -
DR EMBL; Y10768; CAA71741.1; -
DR InterPro: IPR000358; -
DR Pfam: PF00268; ribonuc_red; 1.
DR PROSITE: PS00368; RIBORED_SMALL; 1.
KW Oxidoreductase; DNA replication; Iron.
FT METAL 85 85
FT METAL 116 116 IRON 1 (BY SIMILARITY).
FT METAL 119 119 IRON 1 AND 2 (BY SIMILARITY).
FT METAL 119 119 IRON 1 (BY SIMILARITY).
FT METAL 178 178 IRON 2 (BY SIMILARITY).
FT METAL 212 212 IRON 2 (BY SIMILARITY).
FT METAL 215 215 IRON 2 (BY SIMILARITY).
FT ACT_SITE 123 123 BY SIMILARITY.
FT CONFLICT 8 10 CSR -> RSA (IN REF. 2).
FT CONFLICT 43 43 E -> K (IN REF. 2).
SQ SEQUENCE 337 AA; 39008 MW; BD39919280C65CAE CRC64;

Query Match 19.4%; Score 52; DB 1; Length 337;
Best Local Similarity 31.4%; Pred. No. 20;
Matches 11; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 4 HISSPKYPTQEAQKRAQSRSLGMPGRHLGSSMS 38
DB 27 YVTFPIKPIDWOKYKEASSIWTVEIDLGNDMT 61

RESULT 14
MAPA_RAT
ID MAPA_RAT STANDARD; PRT; 2774 AA.

AC P34926;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2].
GN MAP1A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9235629; PubMed=1379599;
RA Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
RT one messenger RNA.";
RL J. Biol. Chem. 267:16561-16566(1992).
CC -!- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN THE FILAMENTOUS
CC CROSS-BRIDGING BETWEEN MICROTUBULES AND OTHER SKELETAL ELEMENTS.
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
CC APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
CC THEIR MORPHOLOGY.
CC -!- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
CC FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC -!- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B.
CC -!- SIMILARITY: TO MAP1B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M83196; AAB48069.1; -
DR PIR; A43359; A43359.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN 22465 2774 MAP1 LIGHT CHAIN LC2.
FT DOMAIN 309 496 LYS-RICH (BASIC).
FT DOMAIN 336 541 11 X 3 AA REPEATS OF K-K-[DE].
FT REPEAT 336 338 1.
FT REPEAT 415 417 2.
FT REPEAT 420 422 3.
FT REPEAT 424 426 4.
FT REPEAT 427 429 5.
FT REPEAT 431 433 6.
FT REPEAT 436 438 7.
FT REPEAT 440 442 8.
FT REPEAT 444 446 9.
FT REPEAT 449 451 10.
FT REPEAT 539 541 11.
SQ SEQUENCE 2774 AA; 299526 MW; 3DEF74427BA9D7D7 CRC64;

Query Match 19.4%; Score 52; DB 1; Length 2774;
Best Local Similarity 46.2%; Pred. No. 1.8e+02;
Matches 12; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 9 FKYPHTQEAQKRAQSRSLGMPGRHLG 34
DB 678 FYQKHTQEAQKRAQSRSLGMPGRHLG 703

RESULT 15
AQPI_HUMAN

ID AQP1_HUMAN STANDARD; PRT; 269 AA.
AC P29972;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AQUAPORIN-CHIP (WATER CHANNEL PROTEIN FOR RED BLOOD CELLS AND KIDNEY
DE PROXIMAL TUBULE) (AQUAPORIN 1) (AQP-1) (URINE WATER CHANNEL).
GN AQP1 OR CHIP28.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92107900; PubMed=1722319;
RA Preston G.M., Agre P.;
RT "Isolation of the cDNA for erythrocyte integral membrane protein of
RT 28 kilodaltons: member of an ancient channel family.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:11110-11114(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93340184; PubMed=8340403;
RA Moon C., Preston G.M., Griffin C.A., Jabs E.W., Agre P.;
RT "The human aquaporin-CHIP gene. Structure, organization, and
RT chromosomal localization.";
RL J. Biol. Chem. 268:15772-15778(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Retinal pigment epithelium;
RX MEDLINE=96326579; PubMed=8703970;
RA Ruiz A.C., Bok D.;
RT "Characterization of the 3' UTR sequence encoded by the AQP-1 gene in
RT human retinal pigment epithelium.";
RL Biochim. Biophys. Acta 1282:174-178(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=94290349; PubMed=7517253;
RA Li X., Yu H., Koide S.S.;
RT "The water channel gene in human uterus.";
RL Biochem. Mol. Biol. Int. 32:371-377(1994).
RN [5]
RP SEQUENCE OF 1-128 FROM N.A.
RA Dempsey S., Lacy M., Holmes A., Nguyen C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 129-269 FROM N.A.
RA Andrews S., Dubbeide C., Ryan E.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP FUNCTION.
RX MEDLINE=92229472; PubMed=1373524;
RA Preston G.M., Carroll T.P., Guggino W.B., Agre P.;
RT "Appearance of water channels in Xenopus oocytes expressing red cell
RT CHIP28 protein.";
RL Science 256:385-387(1992).
RN [8]
RP TARGET OF MERCURY INHIBITION.
RX MEDLINE=93106996; PubMed=7677994;
RA Preston G.M., Jung J.S., Guggino W.B., Agre P.;
RT "The mercury-sensitive residue at cysteine 189 in the CHIP28 water
RT channel.";
RL J. Biol. Chem. 268:17-20(1993).
RN [9]
RP TOPOLOGY.
RX MEDLINE=94124503; PubMed=7507481;
RA Preston G.M., Jung J.S., Guggino W.B., Agre P.;
RT "Membrane topology of aquaporin CHIP. Analysis of functional epitope-
RT scanning mutants by vectorial proteolysis.";
RL J. Biol. Chem. 269:1668-1673(1994).
RN [10]
RP STRUCTURE BY ELECTRON CRYO-MICROSCOPY.
RX MEDLINE=94313979; PubMed=7518771;

RA Walz T., Smith B.L., Agre P., Engel A.;
RT "The three-dimensional structure of human erythrocyte aquaporin
RT CHIP.";
RL EMBO J. 13:2985-2993(1994).
RN [11]
RP STRUCTURE BY ELECTRON CRYO-MICROSCOPY.
RX MEDLINE=973220502; PubMed=9177353;
RA Walz T., Hirai T., Murata K., Heymann J.B., Mitsuoaka K.,
RA Fujiyoshi Y., Smith B.L., Agre P., Engel A.;
RT "The three-dimensional structure of aquaporin-1.";
RL Nature 387:624-627(1997).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (3.8 ANGSTROMS).
RX MEDLINE=20487015; PubMed=11034202;
RA Murata K., Mitsuoaka K., Hirai T., Walz T., Agre P., Heymann J.B.,
RA Engel A., Fujiyoshi Y.;
RT "Structural determinants of water permeation through aquaporin-1.";
RL Nature 407:599-605(2000).
RN [13]
RP VARIANT BLOOD GROUP COLTON.
RX MEDLINE=94365170; PubMed=7521882;
RA Smith B.L., Preston G.M., Spring F., Anstee D.J., Agre P.;
RT "Human red cell aquaporin CHIP. I. Molecular characterization of ABH
RT and Colton blood group antigens.";
RL J. Clin. Invest. 94:1043-1049(1994).
RN [14]
RP FUNCTION: FORMS A WATER-SPECIFIC CHANNEL THAT PROVIDES THE PLASMA
RC MEMBRANES OF RED CELLS AND KIDNEY PROXIMAL TUBULES WITH HIGH
CC PERMEABILITY TO WATER, THEREBY PERMITTING WATER TO MOVE IN THE
CC DIRECTION OF AN OSMOTIC GRADIENT.
CC [1] SUBUNIT: HOMOTETRAMER.
CC [2] SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC [3] TISSUE SPECIFICITY: EXPRESSED IN A NUMBER OF TISSUES INCLUDING
CC ERYTHROCYTES, RENAL TUBULES, RETINAL PIGMENT EPITHELIUM,
CC HEART LUNG, SKELETAL MUSCLE, KIDNEY AND PANCREAS. WEAKLY
CC EXPRESSED IN BRAIN, PLACENTA AND LIVER.
CC [4] DOMAIN: HB AND HE TOGETHER FORM A TRANSMEMBRANE HELIX. THEY ALSO
CC FUNCTION AS PORE HELICES INVOLVED IN DEFINING SUBSTRATE
CC SPECIFICITY. THE TWO NPA REPEATS ARE SITUATED AT THE END OF HB AND
CC HE AND ARE IMPORTANT FOR THE INTERACTION BETWEEN THESE TWO
CC HELICES.
CC [5] POLYMORPHISM: AQP1 IS RESPONSIBLE FOR THE COLTON BLOOD GROUP
CC SYSTEM. APPROXIMATELY 92% OF CAUCASIANS ARE CO(A+B-) (ALA-45),
CC APPROXIMATELY 8% ARE CO(A+B+), AND ONLY 0.2% ARE CO(A-B+) (VAL-
CC 45). CO(A-B-) WHICH IS VERY RARE, IS DUE TO A COMPLETE ABSENCE OF
CC AQP1.
CC [6] MISCELLANEOUS: PHARMACOLOGICALLY INHIBITED BY SUBMILLIMOLAR
CC CONCENTRATIONS OF Hg2+.
CC [7] SIMILARITY: BELONGS TO THE TRANSMEMBRANE CHANNEL MIP FAMILY.
CC [8] This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC [9] EMBL: M77829; AAA58425.1;
DR EMBL: U41517; AAC50648.1;
DR EMBL: U41518; AAC50649.1;
DR EMBL: S73482; AAB31193.1;
DR EMBL: AC004691; AAC16481.1;
DR EMBL: AC005155; AAC23788.1;
DR PIR: A41616; A41616
DR PDB: 1FQY; 18-OCT-00.
DR MIM: 107776;
DR MIM: 110450;
DR InterPro: IPR000425;
DR Pfam: PF00230; MIP; 1.
DR PRINTS: PR00783; MINTRINSICP.
DR PROSITE: PS00221; MIP; 1.
KW Transport; Transmembrane; Polymorphism; Blood group antigen;
KW Glycoprotein; 3D-structure.

```
FT DOMAIN 1 7 CYTOPLASMIC.
FT TRANSMEM 8 36 HELIX 1.
FT DOMAIN 37 48 EXTRACELLULAR.
FT TRANSMEM 49 66 HELIX 2.
FT DOMAIN 67 70 CYTOPLASMIC.
FT DOMAIN 71 76 IN MEMBRANE.
FT TRANSMEM 77 84 HELIX B.
FT DOMAIN 85 94 CYTOPLASMIC.
FT TRANSMEM 95 115 HELIX 3.
FT DOMAIN 116 136 EXTRACELLULAR.
FT TRANSMEM 137 155 HELIX 4.
FT DOMAIN 156 186 CYTOPLASMIC.
FT TRANSMEM 167 183 HELIX 5.
FT DOMAIN 184 186 EXTRACELLULAR.
FT TRANSMEM 187 192 IN THE MEMBRANE.
FT DOMAIN 193 200 HELIX E.
FT TRANSMEM 201 207 EXTRACELLULAR.
FT TRANSMEM 208 228 HELIX 6.
FT DOMAIN 229 269 CYTOPLASMIC.
FT REPEAT 76 78
FT REPEAT 192 194
FT DOMAIN 159 162 POLY-ARG
FT SITE 189 189 HG(2+)-SENSITIVE RESIDUE.
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .).
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 45 45 A -> V (IN CO(A-B+) ANTIGEN).
FT SEQUENCE 269 AA; 28526 MW; BA204D82FB26352E CRC64;

Query Match 19.2%; Score 51.5; DB 1; Length 269;
Best Local Similarity 26.1%; Pred. No. 18;
Matches 18; Conservative 13; Mismatches 17; Indels 21; Gaps 4;

QY 1 MFLHISP--FKYP--HTQEAQKE-----AQRSLGEMPGRHLSMSLAL- 41
Db 25 VFISIGSALGFKYPVGNNOTAVODNVKVSIAFLGLSTATLAQSVGHISGAHLNPAVTLGLL 84
QY 42 --CLVPLVR 48
Db 85 LSCQISIFR 93
```

Search completed: July 30, 2001, 16:53:32
Job time: 192 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:52:14 ; Search time 26.94 Seconds
(without alignments)
144.206 Million cell updates/sec

Title: US-09-402-713A-7

Perfect score: 268

Sequence: 1 MFLHSSPPKYPHTQPAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62.5	23.3	441	2 E86758	dihydroorotase (EC
2	60.5	22.6	1239	2 T13809	probable disintegr
3	60	22.4	187	2 B81903	hypothetical prote
4	60	22.4	218	2 C81118	hypothetical prote
5	56.5	21.1	1061	2 S75508	acriflavin resista
6	56	20.9	338	2 C75459	probable endonucle
7	55	20.5	1039	2 A41131	lymphocyte-Peyer's
8	54.5	20.3	244	2 A83236	hypothetical prote
9	54.5	20.3	305	2 T47844	hypothetical prote
10	54.5	20.3	461	2 S65665	actin-binding prot
11	54.5	20.3	461	2 S65666	actin-binding prot
12	54	20.1	204	2 T51012	hypothetical prote
13	54	20.1	257	2 G72081	conserved hypothet
14	54	20.1	257	2 F86540	Cn254 hypothetical
15	54	20.1	633	2 B82990	hypothetical prote
16	53.5	20.0	348	2 T14128	NADH dehydrogenase
17	53.5	20.0	446	2 JC4989	gluconate permease
18	53.5	20.0	446	2 E86010	low affinity gluco
19	53.5	20.0	449	2 D72127	hypothetical prote
20	53.5	20.0	451	2 C49898	cellobiose phospho
21	52.5	19.6	216	2 E75446	amino acid ABC tra
22	52.5	19.6	357	2 E83632	hypothetical prote
23	52.5	19.6	441	2 I50515	retinoid X recepto
24	52.5	19.6	493	2 T43465	hypothetical prote
25	52.5	19.6	563	2 T17316	hypothetical prote
26	52	19.4	210	2 H75599	probable acetyltra
27	52	19.4	493	2 F96696	protein FIN21.12 {
28	52	19.4	2774	2 A43359	microtubule-associ
29	51.5	19.2	173	2 D83486	cobinamide kinase

30	51.5	19.2	200	2 H84715	probable phytocyan
31	51.5	19.2	269	2 A41616	erythrocyte integr
32	51.5	19.2	380	2 S75929	sensory transducti
33	51.5	19.2	449	2 C86496	hypothetical prote
34	51.5	19.2	449	2 A81544	hypothetical prote
35	51.5	19.2	470	1 RKMWLX	ribulose-bisphosph
36	51.5	19.2	569	2 T03390	4-coumarate--CoA 1
37	51	19.0	137	2 B72786	hypothetical prote
38	51	19.0	464	2 T45612	N-hydroxycinnamoyl
39	51	19.0	633	2 A75069	dipeptide transpor
40	51	19.0	1038	2 S06046	integrin alpha-4 c
41	51	19.0	2137	2 T05244	hypothetical prote
42	51	19.0	2492	1 A44213	nonstructural poly
43	51	19.0	2492	1 C44213	nonstructural poly
44	50.5	18.8	115	2 G72617	hypothetical prote
45	50.5	18.8	409	1 A47527	transcription fact

ALIGNMENTS

RESULT 1

E86758

dihydroorotase (EC 3.5.2.3) [imported] - Lactococcus lactis subsp. lactis (strain ILI
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C:Accession: E86758

R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh

Genome Res. in press, 2001

A:Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625

A:Accession: E86758

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 <STO>

A:Cross-references: GB:AB005176; NID:g12724023; PIDN:AAK05167.1; GSPDB:GN00146

A:Experimental source: strain ILI403

C:Genetics:

A:Gene: pyrC

C:Superfamily: Bacillus dihydroorotase; Bacillus dihydroorotase homology

C:Keywords: hydrolase

Query Match

Best Local Similarity 23.3%; Score 62.5; DB 2; Length 441;

Matches 17; Conservative 4; Mismatches 16; Indels 3; Gaps 1;

QY 12 PHTQPAQKEAQRSLGEMFGRHLGSSMSLALCLVPLVREGH 51

DB 325 PHTRE---EKNVSLDKAPSGMIGLETSLQLGLTLNVAKGH 361

RESULT 2

T13809

probable disintegrin (EC 3.4.24.-) - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13809

R:Rooken, J.; Pan, D.; Xu, T.; Rubin, G.M.

Science 273, 1227-1231, 1996

A:Title: KUZ, a conserved metalloprotease-disintegrin with two roles during Drosophil

A:Reference number: Z17772; MUID:96355647

A:Accession: T13809

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1239 <ROO>

A:Cross-references: EMBL:U60591; NID:g1531632; PID:g1531633; PIDN:AAK47275.1

C:Genetics:

A:Gene: kuz

A:Cross-references: FlyBase:FBgn0015954

A:Map position: 2

C:Keywords: hydrolase; metalloproteinase

```

Query Match      22.6%; Score 60.5; DB 2; Length 1239;
Best Local Similarity 47.2%; Pred. No. 9.7;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPHTQEAQKQAEQSLGEM----PGRHLGSSMSLA 40
||||| : : : : : | | | | | : : |
Db 341 FKYPH-QKYTEANFAFGAFYDPSTGRRLLGSSANVA 375

RESULT 3
B81903
hypothetical protein NMA1341 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: B81903
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holtroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556
A:Accession: B81903
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-187 <PAR>
A:Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84590.1; PID:g738001
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA1341

Query Match      22.4%; Score 60; DB 2; Length 187;
Best Local Similarity 31.4%; Pred. No. 1.4;
Matches 16; Conservative 11; Mismatches 22; Indels 2; Gaps 1;

QY 1 MFLHSSPKYPHTQEAQKQAEQSLGEMPGRHGSSMSLALCLVPLVRGH 51
: || : : : : : | | : : : | | : : : | |
Db 58 LFLSLRGEIVFPKNETAESEFEKSVHER--RQEGNAGSGRKQLLDLVRGRH 106

RESULT 4
C81118
hypothetical protein NMB1132, NMB1170 [imported] - Neisseria meningitidis (strain MC58 s
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81118; C81133
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizzia, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: C81118
A:Molecule type: DNA
A:Residues: 1-218 <TET>
A:Cross-references: GB:AE002098; NID:g7226363; PIDN:AAF41520.1; PID:g722636
A:Experimental source: serogroup B, strain MC58
A:Accession: H8113
A:Molecule type: DNA
A:Residues: 1-218 <TET>
A:Cross-references: GB:AE002098; NID:g7226401; PIDN:AAF41555.1; PID:g722640
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1132; NMB1170

Query Match      22.4%; Score 60; DB 2; Length 218;
Best Local Similarity 31.4%; Pred. No. 1.6;
Matches 16; Conservative 11; Mismatches 22; Indels 2; Gaps 1;

QY 1 MFLHSSPKYPHTQEAQKQAEQSLGEMPGRHGSSMSLALCLVPLVRGH 51
: || : : : : : | | : : : | | : : : | |

```

```

Db 89 LFLSLRGEIVFPKNETAESEFEKSVHER--RQEGNAGSGRKQLLDLVRGRH 137

RESULT 5
S75508
acriflavin resistance protein acrF - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein slr2131
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S75508
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S75508
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1061 <KAN>
A:Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAAL8069.1; PID:g165
A:Note: the nucleotide sequence was submitted to the EMBL data Library, June 1996
C:Genetics:
A:Gene: acrF
C:Superfamily: acriflavin resistance protein

Query Match      21.1%; Score 56.5; DB 2; Length 1061;
Best Local Similarity 31.0%; Pred. No. 28;
Matches 13; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 5 ISSPKYPHTQEAQKQAEQSLGEMPGRHGSSMSLALCLVPL 46
||| : : : : : | | : : : | | : : : | |
Db 418 ISRFIKVKH-EDPQEAQAEAGELTCATVATSLVMAVPIV 458

RESULT 6
C75459
probable endonuclease III - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: C75459
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: C75459
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <WHI>
A:Cross-references: GB:AE001945; GB:AE000513; NID:g6458642; PIDN:AAF10505.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0928
A:Map position: 1

Query Match      20.9%; Score 56; DB 2; Length 338;
Best Local Similarity 38.1%; Pred. No. 9.3;
Matches 8; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 10 KYPHTQEAQKQAEQSLGEMPG 30
::||| | : : : : | : ||
Db 207 EFPHTPEGHEQARQALALPG 227

RESULT 7
A41131
lymphocyte-Peyer's patch adhesion molecule alpha 4 precursor - mouse
N:Alternate names: integrin alpha-4

```

A:Note: T209.l80
C:Superfamily: Arabidopsis thaliana hypothetical protein At2g44600

Query Match 20.3%; Score 54.5; DB 2; Length 305;
Best Local Similarity 42.5%; Pred. No. 13;
Matches 17; Conservative 7; Mismatches 9; Indels 7; Gaps 3;

QY 16 EAOKEAORS-LGEMPGRH-----LGSSMS-LALCLVPLVR 48
| : | : | : ||| : ||| : |||
Db 210 ESIELRRTPATTPGRKRKIAMGIGKSMGMATCLSPLVR 249

RESULT 10
S65665
actin-binding protein p57 - human
C:Species: Homo sapiens (man)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C:Accession: S65665
R:Suzuki, K.; Nishihata, J.; Arai, Y.; Honma, N.; Yamamoto, K.; Irimura, T.; Toyoshima FBS Lett. 364, 283-288, 1995
A:Title: Molecular cloning of a novel actin-binding protein, p57, with a WD repeat an
A:Reference number: S65614; MUID:95278344
A:Accession: S65665
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-461 <SUZ>
A:CROSS-references: EMBL:D44497; NID:g927648; PIDN:BAA07940.1; PID:g927649
C:Genetics:
A:Gene: GDB:p57
A:CROSS-references: GDB:686252
C:Superfamily: unassigned WD repeat proteins; WD repeat homology
F:77-111/Domain: WD repeat homology <WDR>

Query Match 20.3%; Score 54.5; DB 2; Length 461;
Best Local Similarity 43.8%; Pred. No. 21;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

QY 2 FLHTSPFKYPHTOEAOKEORS LGEMPGRH 33
||| ||| : ||| : ||| : |||
Db 303 FLHYLSMF-----SSKSQGMGYMPKRGL 327

RESULT 11
S65666
actin-binding protein p57 - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 26-May-2000
C:Accession: S65666; S65614
R:Suzuki, K.; Nishihata, J.; Arai, Y.; Honma, N.; Yamamoto, K.; Irimura, T.; Toyoshima FBS Lett. 364, 283-288, 1995
A:Title: Molecular cloning of a novel actin-binding protein, p57, with a WD repeat an
A:Reference number: S65614; MUID:95278344
A:Accession: S65666
A:Molecule type: mRNA
A:Residues: 1-461 <SUZ>
A:CROSS-references: EMBL:D44496; NID:g927646; PIDN:BAA07939.1; PID:d1008527; PID:g927
A:Experimental source: spleen
A:Accession: S65614
A:Molecule type: protein
A:Residues: 21-43; 215-230; 234-245; 334-339; 416-430 <SUZ>
C:Superfamily: unassigned WD repeat proteins; WD repeat homology
C:Keywords: actin binding; leucine zipper
F:77-111/Domain: WD repeat homology <WD1>
F:127-161/Domain: WD repeat homology <WD2>
F:433-454/Region: leucine zipper motif

Query Match 20.3%; Score 54.5; DB 2; Length 461;
Best Local Similarity 43.8%; Pred. No. 21;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

Search completed: July 30, 2001, 16:52:14
Job time: 249 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:50:50 ; Search time 26.85 seconds
(without alignments)
39.110 Million cell updates/sec

Title: US-09-402-713A-7
Perfect score: 268
Sequence: 1 MFLHISPPKYPHTQEAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60.5	22.6	1239	2	US-08-937-931-2
2	60.5	22.6	1239	4	US-09-285-502-2
3	53	19.8	1151	1	US-08-286-889-37
4	53	19.8	1151	1	US-08-485-618-37
5	53	19.8	1151	1	US-08-362-652-37
6	53	19.8	1151	2	US-08-605-672-37
7	53	19.8	1151	2	US-08-482-293A-37
8	53	19.8	1151	2	US-08-943-363-37
9	53	19.8	1151	4	US-09-193-043-37
10	53	19.8	1161	1	US-08-485-618-55
11	53	19.8	1161	1	US-08-362-652-55
12	53	19.8	1161	2	US-08-605-672-55
13	53	19.8	1161	2	US-08-482-293A-55
14	53	19.8	1161	2	US-08-943-363-55
15	53	19.8	1161	4	US-09-193-043-55
16	52.5	19.6	685	3	US-09-031-563-21
17	52.5	19.6	1055	3	US-09-031-563-27
18	52.5	19.6	1315	3	US-09-031-563-2
19	52.5	19.6	1315	3	US-09-031-563-25
20	51.5	19.2	269	1	US-08-447-554-5
21	51.5	19.2	269	1	US-08-468-763-17
22	51.5	19.2	269	1	US-08-448-160-5
23	51.5	19.2	269	2	US-08-393-996A-17
24	51	19.0	1155	1	US-08-286-889-46
25	51	19.0	1155	1	US-08-485-618-46
26	51	19.0	1155	1	US-08-362-652-46
27	51	19.0	1155	2	US-08-605-672-46

28	51	19.0	1155	2	US-08-482-293A-46	Sequence 46, Appl
29	51	19.0	1155	2	US-08-943-363-46	Sequence 46, Appl
30	51	19.0	1155	4	US-09-193-043-46	Sequence 46, Appl
31	51	19.0	1161	1	US-08-173-497-2	Sequence 2, Appli
32	51	19.0	1161	1	US-08-286-889-2	Sequence 2, Appli
33	51	19.0	1161	1	US-08-485-618-2	Sequence 2, Appli
34	51	19.0	1161	1	US-08-485-618-53	Sequence 53, Appl
35	51	19.0	1161	1	US-08-485-618-99	Sequence 99, Appl
36	51	19.0	1161	1	US-08-362-652-2	Sequence 2, Appli
37	51	19.0	1161	1	US-08-362-652-53	Sequence 53, Appl
38	51	19.0	1161	2	US-08-605-672-2	Sequence 2, Appli
39	51	19.0	1161	2	US-08-605-672-53	Sequence 53, Appl
40	51	19.0	1161	2	US-08-605-672-99	Sequence 99, Appl
41	51	19.0	1161	2	US-08-482-293A-2	Sequence 2, Appli
42	51	19.0	1161	2	US-08-482-293A-53	Sequence 53, Appl
43	51	19.0	1161	2	US-08-482-293A-99	Sequence 99, Appl
44	51	19.0	1161	2	US-08-943-363-2	Sequence 2, Appli
45	51	19.0	1161	2	US-08-943-363-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1
US-08-937-931-2
; Sequence 2, Application US/08937931
; Patent No. 5935792
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerald M.
; APPLICANT: Pan, Duojia
; APPLICANT: Rooke, Jenny
; APPLICANT: Yavari, Reza
; APPLICANT: Xu, Tian
; TITLE OF INVENTION: KUZ: A No. 5935792el Family of Metalloproteases
; NUMBER OF INVENTIONS: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,931
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-937-931-2

Query Match 22.6%; Score 60.5; DB 2; Length 1239;
Best Local Similarity 47.2%; Pred. No. 2.9;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;
QY 9 FKYPHTQEAQKQORSIGEM----PGRHLGSSMSLA 40


```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32797
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-618-37

Query Match 19.8%; Score 53; DB 1; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 14 TQEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
      ||||: :| :| :|| :|| :|| :|| :|| :||
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462

RESULT 5
US-08-362-652-37
; Sequence 37, Application US/08362652
; Patent No. 5766850
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,652
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32391
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
; US-08-362-652-37

Query Match 19.8%; Score 53; DB 1; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 14 TQEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
      ||||: :| :| :|| :|| :|| :|| :|| :||
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462

RESULT 6
US-08-605-672-37
; Sequence 37, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605,672
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-605-672-37

Query Match 19.8%; Score 53; DB 2; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 14 TQEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
      ||||: :| :| :|| :|| :|| :|| :|| :||
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462
```

```

> STATE: Illinois
> COUNTRY: United States
> ZIP: 60606-6402
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Floppy disk
> COMPUTER: IBM PC compatible
> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: PatentIn Release #1.0, Version #1.25
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/943,363
> FILING DATE:
> CLASSIFICATION: 530
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 08/173,497
> FILING DATE: 23-DEC-1993
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 08/286,889
> FILING DATE: 5-AUG-1994
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 08/362,652
> FILING DATE: 21-DEC-1994
> ATTORNEY/AGENT INFORMATION:
> NAME: Williams Jr., Joseph A.
> REGISTRATION NUMBER: 38,659
> REFERENCE/DOCKET NUMBER: 27866/32684
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: 312-474-6300
> TELEFAX: 312-474-0448
> TELEX: 25-3856
> INFORMATION FOR SEQ ID NO: 37:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 1151 amino acids
> TYPE: amino acid
> TOPOLOGY: linear
> MOLECULE TYPE: protein
> US-08-943-363-37

Query Match          19.8%; Score 53; DB 2; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels

QY      14   TQEAQKSAQRSLGEMPGRHGLSSMSLALCLVPLVREG 50
        |||||       : | : | : ||| : ||| : | : |
Db       428  TQEARHWRPKS--EVRGTQIGSYFGASLCSDVDVDRGG 462

RESULT      9
US-09-193-043-37
> Sequence 37, Application US/09193043
> Patent No. 6251395
> GENERAL INFORMATION:
> APPLICANT: Gallatin, Michael W.
> APPLICANT: Van der Vieren, Monica
> TITLE OF INVENTION: No. 6251395el Human 2
> FILE REFERENCE: 27866/35004
> CURRENT APPLICATION NUMBER: US/09/193,043
> CURRENT FILING DATE: 1998-11-16
> EARLIER APPLICATION NUMBER: 08/173,497
> EARLIER FILING DATE: 1993-12-23
> EARLIER APPLICATION NUMBER: 08/286,889
> EARLIER FILING DATE: 1994-08-05
> EARLIER APPLICATION NUMBER: 08/362,652
> EARLIER FILING DATE: 1994-12-21
> EARLIER APPLICATION NUMBER: 08/943,363
> EARLIER FILING DATE: 1997-10-03
> NUMBER OF SEQ ID NOS: 114
> SOFTWARE: PatentIn Ver. 2.0
> SEQ ID NO 37
> LENGTH: 1151
> TYPE: PRT
> ORGANISM: Rattus rattus
> OS-09-193-043-37

```


Query Match 19.8%; Score 53; DB 2; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels

QY 14 TQEAQKRAQRSLGEMPGRHILGSSMSIALCLVPLVREG 50
 IIIII : I : I : I : I : I : I : I : I :
 Db 438 TQEAHRWRPKS--EVRGTOIGSYFGASLCSVDVDRDG 472

RESULT 1A

Patent No. 583/418
GENERAL INFORMATION.

```

; ; TITLE OF INVENTION: NO. 5837476el Human 2 Integrin Alpha Subunit
; ; NUMBER OF SEQUENCES: 114
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; ; STREET: 233 South Wacker Drive, 6300 Sear Tower
; ; CITY: Chicago

```

RESULT 13
US-08-482-293A-55

COUNTRY: United
ZTD: 60606-6403

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

OPERATING SYSTEM: PC-D
SOFTWARE: Patent In Rel

```

; SOFTWARE: PATENTCII Release #1.0, VERSION #1.1.23
: CURRENT APPLICATION DATA:

```

APPLCATION NUMBER: 115

; APPLICATION NUMBER: 03/08/343,303
 ; FILING DATE:

CLASSIFICATION: 530

: CLEARIFICATION: 550
 : PRIOR APPLICATION DATA:

APPLICANT AFFILIATION DATA:
APPLICATION NUMBER:

FILING DATE: 23-DEC-

: : PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE: 5-AUG-1964

; PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE: 21-DEC-

ATTORNEY/AGENT INFORMATION

NAME: Williams Jr.,

; ;
REGISTRATION NUMBER:

REFERENCE/DOCKET NUMBER

; TELECOMMUNICATION INFORMATION

TELEPHONE: 312-474-6

TELEFAX: 312-474-0444

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO

; SEQUENCE CHARACTERISTIC

; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-943-363-55

Query Match 19.8%; Score 53; DB 2; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAKQKQSRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
||||: :| | :|| :||| :| |
Db 438 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 472

RESULT 15

US-09-193-043-55
; Sequence 55, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: NO. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-193-043-55

Query Match 19.8%; Score 53; DB 4; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAKQKQSRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
||||: :| | :|| :||| :| |
Db 438 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 472

Search completed: July 30, 2001, 16:50:51
Job time: 166 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:51:40 ; Search time 42.83 seconds
(without alignments)
72.188 Million cell updates/sec

Title: US-09-402-713a-7
Perfect score: 268
Sequence: 1 MFLHTSSPPKYPHTQBAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 50623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	268	100.0	51	AAW79736	Prostate cancer an
2	268	100.0	51	AAW79738	Prostate cancer an
3	77	28.7	16	AAV07887	Human secreted pro
4	63	23.5	514	AAV87343	Human signal pepti
5	63	23.5	602	21 AAB43919	Human cancer assoc
6	60.5	22.6	1239	19 AAW56131	Drosophila melanoga
7	59	22.0	897	20 AAV07092	Colon cancer assoc
8	59	22.0	1141	21 AAB43008	Human ORFX ORP2772
9	56	20.9	11	19 AAW79737	Prostate cancer an
10	55	20.5	184	21 AAB40271	Human ORFX ORF35 p
11	55	20.5	2432	21 AAV85565	Human homologue of

12	54.5	20.3	99	20	AAV29863	Human secreted pro
13	54.5	20.3	100	20	AAV25713	Human secreted pro
14	54.5	20.3	461	17	AAV98341	Human p57 protein.
15	54.5	20.3	461	17	AAV98342	Bovine p57 protein
16	54.5	20.3	598	17	AAV98344	GST/truncated huma
17	54.5	20.3	673	17	AAV98343	GST/human p57 fusi
18	54	20.1	414	21	AAV22216	Arabidopsis thalia
19	54	20.1	414	21	AAV22216	Arabidopsis thalia
20	54	20.1	450	21	AAV22215	Arabidopsis thalia
21	54	20.1	450	21	AAV22215	Arabidopsis thalia
22	54	20.1	453	21	AAV22214	Arabidopsis thalia
23	54	20.1	453	21	AAV22214	Arabidopsis thalia
24	53.5	20.0	112	20	AAV60194	Human endometrium
25	53	19.8	1151	16	AAV78179	Rat alpha-d compos
26	53	19.8	1151	18	AAV23059	Rat beta 2 integri
27	53	19.8	1151	19	AAV72834	Rat alpha-d #2. R
28	53	19.8	1151	19	AAV65101	Rat beta-integrin
29	53	19.8	1151	19	AAV60001	Rat alpha d polype
30	53	19.8	1151	20	AAV73344	Rat alphasad protei
31	53	19.8	1151	21	AAV07371	Rat alpha-d partia
32	53	19.8	1161	16	AAV78169	Rat alpha-d subuni
33	53	19.8	1161	18	AAV23062	Rat beta 2 integri
34	53	19.8	1161	19	AAV72824	Rat alpha-d #1. R
35	53	19.8	1161	19	AAV65104	Rat beta-integrin
36	53	19.8	1161	19	AAV60004	Rat alpha d polype
37	53	19.8	1161	20	AAV73345	Rat alphasad protei
38	53	19.8	1161	21	AAV07374	Rat alpha-d protei
39	52.5	19.6	109	20	AAV48422	Human prostate can
40	52.5	19.6	1055	20	AAV31730	Human fused protei
41	52.5	19.6	1315	20	AAV43265	Human Fused Relate
42	52.5	19.6	1315	20	AAV31725	Human fused protei
43	52.5	19.6	1315	20	AAV31729	Human fused protei
44	52.5	19.6	1315	21	AAV30679	Amino acid sequenc
45	52.5	19.6	1333	20	AAV31727	Human fused cDNA i

ALIGNMENTS

RESULT 1
AAW79736
ID AAW79736 standard; Protein; 51 AA.
XX
AC AAW79736;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) protein variant 1.
XX
KW Prostate cancer antigen protein variant 1; PCA3; prostatic cancer;
PC.
XX
OS Homo sapiens.
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
DR N-PSDB; AAV62427.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
XX prevention and treatment of prostatic cancer
XX
PS Claim 16; Fig 2B-2J; 11lpp; English.

XX The present sequence represents the prostate cancer antigen (PCA3)
 CC protein variant 1 encoded by a PCA3 spliced cDNA sequence comprising
 CC of exons 1, 2, 3, 4a and 4b of the PCA3 gene. The invention claims
 CC for PCA3 cDNA variants and the proteins they encode. The invention
 CC also claims for antibodies against PCA3 protein. The antibodies are
 CC claimed to be useful for detecting PCA3 protein in immunoassay tests,
 CC for diagnosing, assessing and prognosing of prostatic cancer (PC).
 CC Antibodies, optionally coupled to a cytotoxin or radioisotope, and
 CC nucleic acids antisense to PCA3 cDNA are claimed to be useful for
 CC treating PC, while determining elevated levels of PCA3 (as RNA or
 CC protein) is useful for detecting a predisposition to development of
 CC PC, e.g. in prenatal tests. Detecting PCA3 protein allows
 CC differentiation between malignant and benign prostatic disease,
 CC and the level of PCA3 expression allows correlation with the grade of
 CC tumour. PCA3 protein and its fragments are also claimed to be useful
 CC in vaccines for preventing PC; in drug screens for identifying
 CC specific (antagonists (potentially useful therapeutically) and for
 CC studying protein-DNA interactions.
 XX
 SQ Sequence 51 AA;

Query Match 100.0%; Score 268; DB 19; Length 51;
 Best Local Similarity 100.0%; Pred. No. 4.7e-32;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLHSSPFKYPHTQBAQKRAQSLGEMPCRHLSGSSMSLALCLVPLVREGH 51
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 mflhsspfkyphtqbaqkraqslgemprrhlgsmslalcplvrvregh 51

RESULT 2

AAW79738
 ID AAW79738 standard; Protein; 51 AA.

AC AAW79738;

XX 30-DEC-1998 (first entry)

XX Prostate cancer antigen (PCA3) wild-type protein.

XX Prostate cancer antigen protein variant 1; PCA3; prostatic cancer;
 KW PC.

OS Homo sapiens.

XX WO9845420-A1.

XX 15-OCT-1998.

XX 09-APR-1998; 98WO-CA00346.

XX 10-APR-1997; 97US-0041836.

XX (DIAG-) DIAGNOCURE INC.

XX Bussemakers MJG;

XX WPI; 1998-568347/48.

XX N-PSDB; AAV62430.

XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
 PT prevention and treatment of prostatic cancer

XX Claim 16; Fig 5B-5F; 111pp; English.

XX The present sequence represents the prostate cancer antigen (PCA3)
 CC wild-type protein sequence encoded by a PCA3 wild-type cDNA sequence
 CC comprising of exons 1, 2, 3, 4a-4d of the PCA3 gene. The invention
 CC claims for PCA3 cDNA variants and the proteins they encode. The
 CC invention also claims for antibodies against PCA3 protein. The
 CC antibodies are claimed to be useful for detecting PCA3 protein in

CC immunoassay tests, for diagnosing, assessing and prognosing of
 CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
 CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
 CC to be useful for treating PC, while determining elevated levels of
 CC PCA3 (as RNA or protein) is useful for detecting a predisposition to
 CC development of PC, e.g. in prenatal tests. Detecting PCA3 protein
 CC allows differentiation between malignant and benign prostatic disease,
 CC and the level of PCA3 expression allows correlation with the grade of
 CC tumour. PCA3 protein and its fragments are also claimed to be useful
 CC in vaccines for preventing PC; in drug screens for identifying
 CC specific (antagonists (potentially useful therapeutically) and for
 CC studying protein-DNA interactions.
 XX

SQ Sequence 51 AA;

Query Match 100.0%; Score 268; DB 19; Length 51;
 Best Local Similarity 100.0%; Pred. No. 4.7e-32;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLHSSPFKYPHTQBAQKRAQSLGEMPCRHLSGSSMSLALCLVPLVREGH 51
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 mflhsspfkyphtqbaqkraqslgemprrhlgsmslalcplvrvregh 51

RESULT 3

AAV07887

ID AAV07887 standard; Protein; 16 AA.

XX AAV07887;

XX 06-JUL-1999 (first entry)

XX Human secreted protein fragment encoded from gene 36.

XX Human; secreted protein; treatment; prevention; protein therapy; AIDS;
 KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
 KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
 KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
 KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
 KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
 KW arthritis; malignancy; digestive; endocrine; infection.

XX Homo sapiens.

XX WO9918208-A1.

XX 15-APR-1999.

XX 01-OCT-1998; 98WO-US20775.

XX 02-OCT-1997; 97US-0060884.

XX 02-OCT-1997; 97US-0060833.

XX 02-OCT-1997; 97US-0060836.

XX 02-OCT-1997; 97US-0060837.

XX 02-OCT-1997; 97US-0060838.

XX 02-OCT-1997; 97US-0060839.

XX 02-OCT-1997; 97US-0060843.

XX 02-OCT-1997; 97US-0060862.

XX 02-OCT-1997; 97US-0060866.

XX 02-OCT-1997; 97US-0060874.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;
 PI Florence Ka, Greene JM, Janat F, Lafleur DW, Ni J;
 PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;

XX WPI; 1999-264022/22.
 DR N-PSDB; AAX37486.

XX New isolated human genes and the secreted polypeptides they encode
 PT

XX PS Claim 1b; Page 297; 368pp; English.

XX CC This invention describes novel isolated human genes and the secreted

CC proteins they encode. The products of the invention are useful for

CC preventing, treating or ameliorating medical conditions, e.g. by protein

CC or gene therapy. Also pathological conditions can be diagnosed by

CC determining the amount of the new polypeptides in a sample or by

CC determining the presence of mutations in the new polynucleotides.

CC Specific uses are described for each of the 101 polynucleotides, based

CC on which tissues they are most highly expressed in, and include

CC developing products for the diagnosis or treatment of cancer, tumours,

CC neurodegenerative disorders, developmental abnormalities and fetal

CC deficiencies, blood disorders, leukemias, diseases of the immune system,

CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,

CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate

CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,

CC transplant rejection, disorders involving osteoclasts such as

CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,

CC infections and AIDS. The human secreted proteins of the invention are

CC represented in AAY07852-Y07993 and the encoding nucleic acids are

CC represented in AAX37451-X37552.

XX SQ Sequence 16 AA;

Query Match 28.7%; Score 77; DB 20; Length 16;

Best Local Similarity 100.0%; Pred. No. 0.00012; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0;

QY 37 MSLALCLVPLVREGH 51

Db 1 mslalclvplvregh 15

RESULT 4

AAY87343

ID AAY87343 standard; Protein; 514 AA.

XX AC AAY87343;

XX DT 11-MAY-2000 (first entry)

XX DE Human signal peptide containing protein HSPP-120 SEQ ID NO:120.

XX KW Human; signal peptide-containing protein; HSPP; diagnosis; cancer;

KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;

KW antimicrobial; neurotropic; neuroprotective; cardiovascular; hepatotropic;

KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;

KW reproductive disorder; developmental disorder; arteriosclerosis;

KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;

KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;

KW Parkinson's disease; Huntington's diseases; ovulatory defect;

XX OS Homo sapiens.

XX PN WO200000610-A2.

XX PD 06-JAN-2000.

XX PF 25-JUN-1999; 99WO-US14484.

XX PR 26-JUN-1998; 98US-0090762.

XX PR 31-JUL-1998; 98US-0094983.

XX PR 01-OCT-1998; 98US-0102686.

XX PR 11-DEC-1998; 98US-0112129.

XX PA (INCY-) INCYTE PHARM INC.

XX PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;

PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;

PI Bandman O;

XX WPI; 2000-160673/14.

DR N-PSDB; AAZ98228.

XX PT New human signal peptide-containing proteins useful in treatment,

PT prevention and diagnosis of e.g. cancer, inflammation and

PT cardiovascular disease

XX PS Claim 1; Page 239-240; 327pp; English.

XX CC AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the

CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have

CC anticancer, anti-inflammatory, antimicrobial, neurotropic, hepatotropic,

CC neuroprotective, cardiovascular and antiasthmatic activities, and can

CC be used in gene therapy. HSPPs can be used to treat or prevent disorders

CC associated with decreased activity or function of HSPP. Antagonists of

CC HSPP are used to treat or prevent disorders associated with increased

CC activity or function of HSPP. Such diseases include cell proliferation

CC (including cancer), inflammation, cardiovascular, neurological,

CC reproductive or developmental disorders, (e.g. arteriosclerosis,

CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,

CC asthma, Crohn's disease, microbial or other infections, congestive or

CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's

CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP

CC nucleic acids can be used for the recombinant production of HSPP, for

CC detecting HSPP in standard hybridisation and amplification assays (for

CC diagnosis and monitoring), in gene therapy, as antisense,

CC triplex-forming or ribozyme therapeutics, for detecting related sequences

CC or genetic variations, and for chromosomal mapping. HSPP are also used to

CC raise specific antibodies (Ab) and to screen for agonists and

CC antagonists (potential therapeutic agents). Ab are used to diagnose, or

CC monitor, HSPP-related diseases (in usual immunoassays), as therapeutic

CC antagonists, in competitive drug screens, and for purification of HSPP

CC from natural sources.

XX SQ Sequence 514 AA;

Query Match 23.5%; Score 63; DB 21; Length 514;

Best Local Similarity 42.1%; Pred. No. 0.92; Mismatches 13; Indels 2; Gaps 2;

Matches 16; Conservative 7;

QY 12 PHTQEAQKQKSL-GEMPGRHLSMSLAL-CLVPLV 47

Db 471 praedlaedsgslygrapgrhtwslilaalclvpll 508

RESULT 5

AAB43919

ID AAB43919 standard; Protein; 602 AA.

XX AC AAB43919;

XX DT 08-FEB-2001 (first entry)

XX DE Human cancer associated protein sequence SEQ ID NO:1364.

XX KW Human; cancer associated gene; cancer antigen; detection; cancer;

KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;

KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;

KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;

KW dermatological; neuroprotective; thrombolytic; coagulant; neurotropic;

KW vasotrophic; antipsoriatic; antiangiogenic; gene therapy; inflammation;

KW immune disorder; haematopoietic cell disorder; autoimmune disorder;

KW allergic reaction; graft versus host disease; organ rejection;

KW haemostatic; thrombolytic; cardiovascular disorder; infection;

KW neurological disease; drug screening.

XX OS Homo sapiens.

XX PN WO2000055350-A1.

XX PD 21-SEP-2000.

```
XX PF 08-MAR-2000; 2000WO-US05882.
XX XX
XX PR 12-MAR-1999; 99US-0124270.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX XX
XX PI Rosen CA, Ruben SM;
XX XX
XX DR WPI: 2000-587533/55.
XX DR N-PSDB; AAC78128.
XX XX
XX PT Novel isolated nucleic acids comprising sequences encoding peptides
XX PT useful for treating or diagnosing e.g. cancer -
XX XX
XX PS Claim 11; Page 2024-2026; 2352pp; English.
XX XX
XX CC AAC77607 to AAC78448 encode the human cancer associated proteins given
XX CC in AAB43398 to AAB44239. The proteins can have activities based on the
XX CC tissues and cells the genes are expressed in. Example of activities
XX CC include: cytostatic; proliferative; vulnerable; immunomodulator;
XX CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
XX CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
XX CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
XX CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
XX CC polynucleotides and polypeptides can be used for preventing, treating or
XX CC ameliorating medical conditions and diagnosing pathological conditions.
XX CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
XX CC the present invention may be used to treat immune disorders by activating
XX CC or inhibiting the proliferation, differentiation or mobilisation of
XX CC immune cells, to treat disorders of haematopoietic cells, autoimmune
XX CC disorders, allergic reactions, graft versus host disease and organ
XX CC rejection, modulate haemostatic or thrombolytic activity, modulate
XX CC inflammation, cancers, cardiovascular disorders, neurological disease and
XX CC bacterial or viral infections. The peptides, nucleotides, antibodies,
XX CC agonists and antagonists may be also be used in drug screens. AAC78449 to
XX CC AAC78457 and AAB44240 represent sequences used in the exemplification of
XX CC the present invention.
XX XX
XX SQ Sequence 602 AA;

Query Match 23.5%; Score 63; DB 21; Length 602;
Best Local Similarity 42.1%; Pred. No. 1.1;
Matches 16; Conservative 7; Mismatches 13; Indels 2; Gaps 2;

QY 12 PHTQEAQKAQRL-GEMPGHLSMSLAL-CLVPLV 47
| : : : : | | | | | : | | | | |
Db 559 praediaedsgsslyrapgrhtwslllaalacvlpll 596

RESULT 6
AAW56131
ID AAW56131 standard; Protein; 1239 AA.
XX AC
XX AC AAW56131;
XX XX
XX DT 17-AUG-1998 (first entry)
XX XX
XX DE Drosophila melanogaster KUZ protein.
XX XX
XX KW kuzbanian; kuz; neurogenic; KUZ protein; neuronal partitioning;
XX KW development; NOTCH protein processing; regulation; cell function;
XX KW signal transduction pathways; screening; receptor binding;
XX KW metalloprotease.
XX XX
XX OS Drosophila melanogaster.
XX XX
XX PN WO9808933-A1.
XX XX
XX PD 05-MAR-1998.
XX XX
XX PF 27-AUG-1997; 97WO-US15099.
```

```
XX XX
XX PR 23-JUL-1997; 97US-0019390.
XX PR 29-AUG-1996; 96US-0019390.
XX XX
XX PA (REGC ) UNIV CALIFORNIA.
XX PA (UYIA ) UNIV YALE.
XX XX
XX PI Pan D, Rooke J, Rubin GM, Xu T, Yavari R;
XX XX
XX DR WPI: 1998-179428/16.
XX DR N-PSDB; AAV22647.
XX XX
XX PT New KUZ polypeptides, members of the ADAM family of metalloprotease
XX PT - useful in neural partitioning and development
XX XX
XX PS Claim 1; Pages 31-34; 58pp; English.
XX XX
XX CC The sequence is that of encoded by the kuzbanian gene, a
XX CC neurogenic gene. The KUZ protein is involved in neuronal
XX CC partitioning and development. It is also involved in processing
XX CC of the NOTCH protein by cleaving the C-terminal portion. The
XX CC KUZ protein can be used to screen for compounds that alter
XX CC binding of KUZ to its receptor or the cleavage of the NOTCH
XX CC protein, hence acting to regulate NOTCH signal transduction
XX CC pathways and regulate cell functions.
XX XX
XX SQ Sequence 1239 AA;

Query Match 22.6%; Score 60.5; DB 19; Length 1239;
Best Local Similarity 47.2%; Pred. No. 6.2;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPHTQEAQKAQRLGEM----PGRHLGSSMSLA 40
| | | | | : | | | : | | | | | : |
Db 341 fkyph-qktykeanfaegafydpstgrigssanva 375

RESULT 7
AAV07092
ID AAV07092 standard; Protein; 897 AA.
XX XX
XX AC AAV07092;
XX XX
XX DT 02-JUL-1999 (first entry)
XX XX
XX DE Colon cancer associated antigen precursor sequence.
XX XX
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
XX KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX KW prostate cancer.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO9904265-A2.
XX XX
XX PD 28-JAN-1999.
XX XX
XX PF 15-JUL-1998; 98WO-US14679.
XX XX
XX PR 22-JUN-1998; 98US-0102322.
XX PR 17-JUL-1997; 97US-0896164.
XX PR 10-OCT-1997; 97US-0061599.
XX PR 10-OCT-1997; 97US-0061765.
XX PR 10-OCT-1997; 97US-0948705.
XX PR 11-OCT-1997; 97GB-0021697.
XX XX
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX XX
XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
XX PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
XX PI Tureci O;
XX XX
```


FT	Region	1776..1778	
FT	/note=	"present or absent depending upon the allele	
FT		from which the protein is translated"	
XX			
PN	WO963080-A1.		
XX			
PD	09-DEC-1999.		
XX			
PF	02-JUN-1999;	99WO-EP03848.	
XX			
PR	03-JUN-1998;	98CB-0011962.	
XX			
PA	(JANC) JANSSEN PHARM NV.		
XX			
PI	Luyten WHML, De Raeymaeker MC, Geysen JJCH, Bogaert TAOE;		
XX	Maerten LJS, Verhasselt P, Van De Craen M;		
XX			
DR	WPI; 2000-116370/10.		
DR	N-PSDB; AAA07836.		
XX			
PT	Novel proteins and nucleic acids e.g. for treating neurodegeneration		
XX			
PS	Claim 95; Fig 1d; 146pp; English.		
XX			
CC	The invention provides vertebrate (human) protein homologue of a UNC-53		
CC	protein of <i>Caenorhabditis elegans</i> . The UNC-53 binds to microtubules or		
CC	their plus ends. The UNC-53 sequences are used to promote neural		
CC	regeneration, revascularization and wound healing; also for treating		
CC	neurodegenerative disease, acute traumatic injury, fibrotic disease and		
CC	autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53		
CC	polynucleotides can be used for recombinant production of the proteins,		
CC	as a source of probes for detecting allelic variants and polymorphisms,		
CC	for sequencing genomic DNA and for detecting UNC-53 expression; and as		
CC	source of therapeutic antisense sequences. Cells that express the		
CC	protein are used to identify regulators of cell shape, growth, motility		
CC	and migration. They can also be used to identify proteins that are		
CC	involved in signal transduction pathways also involving UNC-53, and to		
CC	identify compounds that alter attachment of UNC-53 to microtubules. A		
CC	target gene coupled to a UNC-53 encoding sequence may be used to deliver		
CC	the target gene to a cellular microtubule or its plus ends. The present		
CC	sequence represents the amino acid sequence of the second human homologue		
CC	of UNC-53, designated hs-UNC-53/2.		
XX			
XX	Sequence	2432 AA;	
XX			

```

Query Match.          20.5%; Score 55; DB 21; Length 2432;
Best Local Similarity 34.0%; Pred. No. 90;
Matches 16; Conservative 3; Mismatches 16; Indels 12; Gaps 1;

QY      6  SSPFKYPTQEAQKEAQRSLG-----EMPCRHLGSSMSLA 40
      | | | | | | | | | | | | | | | | | | | |
Db      571  smpkqspapspapspkqegrsrgklsqllpqkqkpldldarhssssssla 617

```

RESULT	12	
AAV29863		
ID	AAV29863 standard; Protein; 99 AA.	
XX		
XX		
XX	AAV29863;	
XX		
XX		
DT	17-NOV-1999 (first entry)	
DT		
DE	Human secreted protein clone gn82_6.	
DE		
XX		
XX	Human; secreted protein; biological activity; nutritional; cytokine;	
KW	cell proliferation; differentiation; immune stimulating; vaccine;	
KW	haematopoiesis regulation; tissue growth; haemostatic; thrombolytic;	
KW	anti-inflammatory; tumour inhibition.	
XX		
XX		
OS	Homo sapiens.	
XX		
XX		
PN	WO9946287-A1.	

XX	16-SEP-1999.	
PD		
XX		
XX	11-MAR-1999; 99WO-US05243.	
PF		
XX		
XX	11-MAR-1998; 98US-0077521.	
FR		
PR	14-MAY-1998; 98US-0079124.	
PR	10-MAR-1999; 99US-0266105.	
XX		
PA	(GEMY) GENETICS INST INC.	
XX		
XX	Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;	
PI	Merberg D, Treacy M, Agostino MU, Steininger RJ;	
XX		
DR	WPI; 1999-551362/46.	
DR	N-PSDB; AAZ21095.	
XX		
XX	Polynucleotides encoding secreted human proteins, derived from human	
PT	fetal brain, human adult blood, human adult bladder, or human adult	
PT	neural tissue cDNA libraries.	
XX		
XX	Claim 15; Page 103; 118pp; English.	
PS		
XX		
CC	AAZ21093 to AAZ21102 encode new human secreted proteins and AAY29861 to	
CC	AAZ29873 represent the secreted proteins encoded by the polynucleotide	
CC	sequences. AAZ21103 to AAZ21112 represent probes for the secreted	
CC	proteins. The polynucleotides and proteins are predicted to have	
CC	biological activities which would make them suitable for treating,	
CC	preventing or ameliorating medical conditions in humans and animals,	
CC	although no supporting data is given. Suggested activities include	
CC	nutritional activity, cytokine and cell proliferation/differentiation	
CC	activity, immune stimulating (e.g. as vaccines) or suppressing activity,	
CC	haematopoiesis regulating activity, tissue growth activity,	
CC	activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	and thrombolytic activity, receptor/ligand activity, anti-inflammatory	
CC	activity, cadherin/tumour invasion suppressor activity, and tumour	
CC	inhibition activity. The polynucleotides and proteins can also be used	
CC	as nutritional sources or supplements. Such uses include use as a protein	
CC	or amino acid supplement, use as a carbon source, use as a nitrogen	
CC	source and use as a source of carbohydrate. They may also have utility	
CC	in compositions used for bone, cartilage, tendon, ligament, and/or nerve	
CC	tissue growth or regeneration, as well as for wound healing and tissue	
CC	repair and replacement, and in the treatment of burns, incisions and	
CC	ulcers. The proteins which induce cartilage and/or bone growth in	
CC	circumstances where bone is not normally formed, have application in	
CC	the healing of bone fractures and cartilage damage or defects in humans	
CC	and other animals.	
XX		
XX	Sequence 99 AA:	

```

Query Match      20.3%; Score 54.5; DB 20; Length 99;
Best Local Similarity 36.0%; Pred. No. 2.2;
Matches 18; Conservative 7; Mismatches 20; Indels 5; Gaps 2;

Oy   1  MFLHTSSPFKYPHTQEAQKEAQRSLGEMPGRHGLGSSMSLALCLVPLVREG 50
       :||| | | | | | : : | | | :||| :|| | | | | : |
Db    16  lfltsrtfvip--vhavksaqyleeivkhelgsqsl---lspveepg 60

RESULT 13
AAAY25713
ID AAY25713 standard; protein: 100 AA.
```

DT 04-OCT-1999 (first entry)

DE Human secreted protein encoded from gene 3.

Secreted protein; human; treatment; diagnosis; therapy; cancer; tumour;
KW neurodegenerative disorder; developmental abnormality; blood disorder;
KW fetal deficiency; blood disorder; leukemia; immune system inflammation;
KW


```

FT Peptide /note= "WD40 repeat fragment"
FT 127..161
FT /note= "WD40 repeat fragment"
FT 172..203
FT /note= "WD40 repeat fragment"
FT 215..252
FT /note= "WD40 repeat fragment"
FT 265..298
FT /note= "WD40 repeat fragment"
FT 433..461
FT /note= "Leucine rich C terminal fragment"

```

XX JP08119996-A.

XX 14-MAY-1996.

XX 21-OCT-1994; 94JP-0282743.

XX 21-OCT-1994; 94JP-0282743.

XX (NISB) JAPAN TOBACCO INC.

XX WPI; 1996-283507/29.

XX N-PSDB; AAT30361.

XX A new protein, p57, comprising WD40 repeat region - used for
development of anti-cancer and anti-HIV agents

XX Example 2; Page 27-29; 51pp; Japanese.

XX This sequence represents the bovine p57 protein. p57 contains a leucine
rich C-terminal peptide which comprises a leucine residue after each
7 amino acids and a WD40 repeated structure region containing five
WD40 regions. p57 forms a dimer. It can be used in the development
of an anti-cancer agent and an anti-HIV agent. It can also combine
with actin in the control of cell movement.

XX Sequence 461 AA;

Query Match 20.3%; Score 54.5; DB 17; Length 461;

Best Local Similarity 43.8%; Pred. No. 14;

Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

OY 2 FLHISSPFKYPHTQEAOKAORSILGEMPGRHLL 33

Db 303 flhylsmf-----sskesqrgmgypkrgl 327

Search completed: July 30, 2001, 16:51:41

Job time: 216 sec

THIS PAGE BLANK (USPTO)